Joinforcs: Journal of informatics and information system
Volume 1, (3) 2026, p. 93-98
E-ISSN : XXXX-XXXX

Analysis of the Certainty Factor Method for Depression Severity Classification in University Students

Likel Dorisian P[a,1,*], Achmad Nur Sholeh[b,2]
[a]Department of Information System, STMIK WICIDA Samarinda, Indonesia
[b]Department of Informatic Engineering, Universitas Pamulang, Indonesia
[1]2341047@wicida.ac.id, [2]dosen01531@unpam.ac.id
*corresponding author

Article Info

Article History:

Received Feb 22, 2026
Revised Mar 19, 2026
Accepted April 4, 2026

Keywords :
Certainty Factor, Confusion Matrix, Depression Severity, Expert System, Mental Health

Abstract

Depression among university students requires accessible preliminary screening, yet the reliability of simplified rule-based models for representing multiple severity levels remains unclear. This study evaluates the performance of a Certainty Factor (CF) model for classifying depression severity using a publicly available secondary dataset containing 783 student records. After excluding four records with missing depression_severity labels, 779 records were analyzed. The model used three binary evidence variables depressiveness, suicidal ideation, and sleepiness and combined pre-specified Measure of Belief and Measure of Disbelief weights using the CF combination rule. The resulting CF scores were mapped to five predefined classes: None-minimal, Mild, Moderate, Moderately severe, and Severe. Model predictions were compared with questionnaire-derived depression-severity labels using a 5 × 5 multiclass confusion matrix and overall accuracy. The model correctly classified 315 of 779 records, producing an accuracy of 40.44% and an error rate of 59.56%. The evaluation also revealed that the model generated no Moderate predictions because none of the attainable CF scores fell within the Moderate interval. In addition, many Mild cases were classified as None-minimal because the three evidence variables did not represent the broader symptom profile captured by the reference questionnaire. These findings indicate that the evaluated CF configuration is insufficient for reliable five-class depression-severity classification. Future research should incorporate all PHQ-9 indicators, derive CF weights through a documented expert-elicitation process, justify missing-data handling, and validate the model on an independent dataset.

1. Introduction

Depression is a mental-health concern that can affect students' wellbeing, cognitive functioning, and academic engagement. Recent expert-system studies have applied Certainty Factor (CF) reasoning to mental-health screening and depression-related decision support, illustrating the need for accessible preliminary assessment tools [1]. Early screening is valuable when access to professional support is constrained by limited availability, cost, stigma, or low awareness. A computer-based expert system can provide an initial, non-clinical indication from stated symptoms and can support referral to qualified professionals; it does not replace a psychologist or psychiatrist [2].

In addition to diagnostic accuracy, digital health applications should also provide good usability to ensure that users can effectively interact with the system during preliminary health screening. Previous studies on mobile health applications have shown that usability and user experience are essential factors influencing the effectiveness and acceptance of health technologies [3].

Joinforcs: Journal of informatics and information system
Volume 1, (3) 2026, p. 93-98
E-ISSN : XXXX-XXXX

The assessment of depressive symptoms contains uncertainty because an individual symptom may have different meanings and reported intensity may vary. The CF method represents this uncertainty by combining a Measure of Belief (MB) and a Measure of Disbelief (MD) for evidence associated with a hypothesis. This characteristic has made CF suitable for rule-based preliminary screening systems [4]. A limitation of many CF-based applications is that they emphasize application implementation or binary diagnostic output rather than evaluating whether a small set of binary inputs can reproduce a multi-level severity label. In this study, the reference labels correspond to the five PHQ-9 severity bands: None-minimal, Mild, Moderate, Moderately severe, and Severe [5]. Accordingly, this study evaluates the algorithmic accuracy of a pre-specified CF model for classifying depression severity in university-student records. The evaluation compares the predicted severity class from three binary symptom variables with the questionnaire-derived depression-severity label using a multiclass confusion matrix. The study is an algorithmic simulation and is not intended to establish a clinical diagnosis.

Prior work demonstrates the practical use of CF in mental-health expert systems. [6] and [7] combined Forward Chaining and CF to guide diagnostic reasoning. [8], [9], [10], and [11] developed CF-based systems for depression or broader mental-health conditions. These studies show that CF can be implemented in rule-based applications, but their primary contribution is commonly system development rather than large-sample, multiclass evaluation. Other studies broaden the target population or psychological construct. [12] applied CF to bipolar disorder; [13] and [14] addressed anxiety; [15] incorporated DASS-42 indicators for stress; [15] focused on adolescents; and [16] studied people aged 18–23 years. These works demonstrate the flexibility of CF, while also showing that the quality of the result depends on the completeness of symptoms, the knowledge-base weights, and the evaluation design. The present study differs by evaluating a five-class depression-severity output against 779 questionnaire-derived reference labels. This evaluation design makes it possible to identify class-level failure patterns, particularly whether the CF score combinations can reach every predefined severity interval.

A recent systematic review of expert systems for mental-health disorders in college students also reported the widespread use of Certainty Factor and emphasized that model performance depends on the selected symptoms, knowledge-base weights, and evaluation procedure [16].

2. Research Methodology

The methodology maps a secondary student dataset into a rule-based CF simulation, converts the combined CF output to a depression-severity class, and evaluates the predicted class against the corresponding reference label. The steps comprise data preprocessing, CF calculation, score-to-class mapping, and multiclass performance evaluation.

A. Dataset and Data Preprocessing

The study used the publicly available Depression and Anxiety dataset, which contains 783 records and 19 variables [6]. This study analyzes only the depression-severity task; it does not evaluate anxiety severity. Four records with depression_severity coded as NA were excluded. The remaining 779 records were retained. The raw label none (15 records) was standardized to None-minimal. Among the retained records, eight NA values occurred in sleepiness and were coded as FALSE for the binary CF simulation; depressiveness and suicidal contained no missing values after the target-label exclusion. This rule preserved 779 records and is treated as a limitation when interpreting the results.

The three CF evidence variables are depressiveness, suicidal, and sleepiness. The phq_score and depression_severity variables are used only as reference information for evaluation, so they are not used to calculate the CF prediction. Although the source dataset also includes anxiety-related variables, they are outside the scope of this depression-only analysis. The eight missing sleepiness values in the retained records were coded as FALSE solely to preserve the stated 779-record evaluation set, show in Table 1.

Joinforcs: Journal of informatics and information system
Volume 1, (3) 2026, p. 93-98
E-ISSN : XXXX-XXXX

Table 1. Variables Used in the Depression Severity Analysis

Variable	Data type	Role	Preprocessing / use
depressiveness	Boolean	CF evidence	TRUE = present; FALSE = absent.
suicidal	Boolean	CF evidence	TRUE = present; FALSE = absent.
sleepiness	Boolean	CF evidence	TRUE = present; FALSE = absent; 8 NA values coded as FALSE.
phq_score	Integer	Reference score	Not used as CF input; associated with the reference severity label.
depression_severity	Categorical	Ground-truth label	4 NA records excluded; none standardized to None-minimal.
age, gender, school_year	Mixed	Demographics	Descriptive variables; not used as CF inputs.

B. Certainty Factor Formulation and Severity Mapping

For each evidence variable E associated with a depression-severity hypothesis H, the CF value is calculated from the assigned MB and MD values. The weights are pre-specified simulation parameters rather than weights estimated from the dataset.

$$CF(H, E) = MB(H, E) - MD(H, E)$$

When more than one positive evidence variable is present, the individual CF values are combined iteratively using the positive-CF combination rule.

$$CFcombine(CF1, CF2) = CF1 + CF2 \times (1 - CF1)$$

The final combined CF score is then mapped to a predicted severity class using the predefined intervals in Table 2. The intervals are a modeling rule used to enable comparison between the continuous CF output and the categorical reference labels.

Table 2. Predefined Mapping of CF Scores to Depression Severity Classes

CF score range	Predicted severity class
$0.00 \leq CF \leq 0.20$	None-minimal
$0.20 < CF \leq 0.40$	Mild
$0.40 < CF \leq 0.60$	Moderate
$0.60 < CF \leq 0.80$	Moderately severe
$0.80 < CF \leq 1.00$	Severe

The mapping in Table 2 is a predefined classification rule. It is not a property of the source dataset and is not clinically calibrated in this study.

C. Evaluation of Classification Performance

The predicted class for each record was compared with the standardized depression_severity label in a 5 × 5 multiclass confusion matrix. Overall accuracy was calculated from the diagonal of the matrix divided by the total number of evaluated records.

$$Accuracy = (Correct\ predictions / Total\ records) \times 100\%$$

Joinforcs: Journal of informatics and information system
Volume 1, (3) 2026, p. 93-98
E-ISSN : XXXX-XXXX

4. Results and Discussion

A. Assigned CF Values for the Depression Severity Simulation

Table 3 presents the MB and MD values used in the depression-severity simulation. Because this study uses a secondary dataset and does not report a formal expert-elicitation procedure, the values should be interpreted as modeling assumptions used to assess CF behavior, not as validated clinical weights. The configuration assigns the highest individual CF value to suicidal ideation, followed by depressiveness and sleepiness.

Table 3. CF Weights Used in the Depression Severity Simulation

Code	Evidence variable	MB	MD	CF (MB - MD)
G01	Depressiveness	0.80	0.10	0.70
G02	Suicidal ideation	0.90	0.05	0.85
G03	Sleepiness	0.50	0.20	0.30

B. Case Calculation Simulation

To illustrate the rule calculation, consider a record with depressiveness (G01) and sleepiness (G03) coded TRUE and suicidal (G02) coded FALSE. Only G01 and G03 contribute to the CF combination.

$$1.\ Evidence\ G01{:}\ CF1 = 0.70$$
$$2.\ Evidence\ G03{:}\ CF2 = 0.30$$
$$3.\ CFcombine = CF1 + CF2 \times (1 - CF1)$$
$$CFcombine = 0.70 + 0.30 \times (1 - 0.70)$$
$$CFcombine = 0.70 + 0.30 \times 0.30$$
$$CFcombine = 0.70 + 0.09 = 0.79\ (79\%)$$

According to Table 2, a score of 0.79 falls in the interval 0.60 < CF ≤ 0.80 and is therefore classified as Moderately severe.

C. Multiclass Confusion Matrix Evaluation

Table 4. Multiclass Confusion Matrix for Depression Severity Predictions

Actual class \ Predicted class	None-min.	Mild	Moderate	Moderately severe	Severe
None-minimal	226	13	0	0	2
Mild	270	60	0	0	14
Moderate	0	0	0	117	23
Moderately severe	0	0	0	21	21
Severe	0	0	0	4	8

The model correctly classifies 315 of 779 records, resulting in an accuracy of 40.44%. The remaining 464 records are misclassified, corresponding to an error rate of 59.56%. The results indicate that the current CF configuration is not adequate for five-class depression-severity classification. The performance limitation is attributable to the restricted binary evidence set, the pre-specified weights, and the predefined score-to-class mapping.

First, the model has a prediction blind spot for the Moderate class. Under the specified evidence weights, the attainable CF scores are 0.00, 0.30, 0.70, 0.79, 0.85, 0.895, 0.955, and 0.9685. None fall in the Moderate interval (0.40 < CF ≤ 0.60), so the model cannot predict the Moderate class under this configuration. Second, the evidence variables under-represent the symptom set needed to distinguish mild and intermediate depression severity. For example, 270 records labeled Mild are predicted as None-minimal because no positive value is present for depressiveness, suicidal, or sleepiness. Symptoms not represented by these three variables, such as reduced interest, fatigue, appetite change, or concentration difficulty, can still contribute to the questionnaire-derived reference label [5].

The result should therefore be interpreted as an algorithmic simulation rather than a clinical diagnostic validation. The limited three-variable architecture cannot reproduce the full five-level depression-severity spectrum represented by the PHQ-9 [5].

5. Conclusion

This study evaluated a pre-specified Certainty Factor model for classifying depression severity in university-student records. Using 779 valid records and a five-class multiclass confusion matrix, the model achieved an accuracy of 40.44% and an error rate of 59.56%. The result indicates that the present configuration is insufficient for reliable five-class depression-severity classification.

The evaluation identifies two central limitations. First, the selected weights and mapping thresholds create a Moderate-class blind spot because no attainable CF score falls in its interval. Second, three binary evidence variables do not represent the full symptom profile behind the questionnaire-derived severity label, leading to substantial misclassification of Mild cases.

Future work should incorporate all PHQ-9 indicators, obtain CF weights through a documented expert-elicitation process, define and justify missing-data handling, and validate the model using an independent dataset. Comparative testing against data-driven or hybrid methods may also clarify whether the CF rule base can be improved. Until such validation is completed, the model should be positioned only as a research simulation, not as a stand-alone clinical diagnostic tool.

REFERENCES

[1] M. T. Anindita, Y. Pristyanto, H. Sismoro, A. Nurmasani, and A. F. Nugraha, “Diagnose of Mental Illness Using Forward Chaining and Certainty Factor,” Jurnal Techno Nusa Mandiri, vol. 20, no. 2, pp. 63–70, 2023, doi: https://doi.org/10.33480/techno.v20i2.4330.

[2] R. Maliyah and G. Z. Muflih, “Sistem Pakar Diagnosa Depresi Mahasiswa Tingkat Akhir Menggunakan Metode Certainty Factor,” Journal of Science and Social Research, vol. 8, no. 4, pp. 5339–5348, 2025, doi: https://doi.org/10.54314/jssr.v8i4.5085.

[3] Sobri, M., Ijab, M. T., Mat Nayan, N and Tondas., A. E. 2023. The Affordances and Usability Evaluation for HeartM 3.0: A Mobile Heart Monitoring Application. In Advances in Visual Informatics: 8th International Visual Informatics Conference, IVIC 2023, Selangor, Malaysia, November 15–17, 2023, Proceedings. Springer-Verlag, Berlin, Heidelberg, 209–222. https://doi.org/10.1007/978-981-99-7339-2_19

[4] B. A. Nugroho, A. Sulistyohati, and A. N. Arfa, “Implementasi Metode Certainty Factor pada Sistem Pakar Diagnosa Gangguan Kesehatan Mental,” Jurnal Riset dan Aplikasi Mahasiswa Informatika, vol. 6, no. 2, pp. 347–356, 2025, doi: https://doi.org/10.30998/jrami.v6i02.12852.

[5] D. I. Sipata and I. Susilawati, “Sistem Pakar Diagnosa Gangguan Kecemasan Menggunakan Metode Certainty Factor,” JEKIN: Jurnal Teknik Informatika, vol. 5, no. 2, pp. 1007–1021, 2025, doi: https://doi.org/10.58794/jekin.v5i2.1593.

[6] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, “The PHQ-9: Validity of a Brief Depression Severity Measure,” Journal of General Internal Medicine, vol. 16, no. 9, pp. 606–613, 2001, doi: https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

[7] S. Ahmad, “Depression and anxiety data,” Kaggle Dataset. [Online]. Available: https://www.kaggle.com/datasets/shahzadahmad0402/depression-and-anxiety-data. [Accessed: Jun. 25, 2026].

[8] F. Ramadhan, Yuhandri, and G. W. Nurcahyo, “Penerapan Forward Chaining dan Metode Certainty Factor dalam Merancang Sistem Pakar Diagnosa Gangguan Kepribadian,” Jurnal KomtekInfo, vol. 11, no. 4, pp. 213–221, 2024, doi: https://doi.org/10.35134/komtekinfo.v11i4.548.

[9] H. Darwis, P. A. Rahmasari, and Irawati, “Sistem Pakar Mendiagnosis Penyakit Gangguan Mental dengan Metode Certainty Factor Berbasis Android,” Buletin Sistem Informasi dan Teknologi Islam, vol. 6, no. 2, pp. 47–57, 2025, doi: https://doi.org/10.33096/busiti.v5i4.2391.

[10] A. Muhammad, B. Hendrik, and R. Iswara, "Expert System Application for Diagnosing Bipolar Disorder with Certainty Factor Method Based on Web and Android," Journal of Physics: Conference Series, vol. 1339, no. 1, Art. no. 012020, 2019, doi: https://doi.org/10.1088/1742-6596/1339/1/012020.

[11] W. Pradani, S. Hani, S. Rahmawati, A. Jamal, and A. Arniaty Arsyad, "A Knowledge-Based Stress Diagnosis System Using the Certainty Factor Approach and DASS-42 Psychological Indicators," in Proceedings of the Seminar Nasional Penelitian dan Pemberdayaan Masyarakat (SENDAMAS), vol. 5, no. 1, pp. 89–95, 2025, doi: https://doi.org/10.36722/psn.v5i1.5018.

[12] N. I. Siahaan, "Sistem Pakar untuk Diagnosis Awal Indikasi Gangguan Kecemasan Menggunakan Metode Certainty Factor," undergraduate thesis, Universitas Islam Negeri Sultan Syarif Kasim Riau, Riau, Indonesia, 2025.

[13] A. Wildan and R. Sari, "Metode Certainty Factor pada Sistem Pakar Identifikasi Penyakit Mental," Journal of Informatics and Information Security (JIFORTY), vol. 3, no. 1, pp. 125–136, 2022, doi: https://doi.org/10.31599/1n4z9070.

[14] N. W. Nurpanto and P. Kasih, "Sistem Cerdas Sistem Bantu Pengenalan Awal Kesehatan Mental pada Remaja Sekolah Menengah Pertama dengan Forward Chaining dan Certainty Factor," Nusantara of Engineering, vol. 8, no. 1, pp. 27–35, 2025, doi: https://doi.org/10.29407/noe.v8i01.23346.

[15] A. Tasari, E. A. M. Simanjutak, G. Christian, and R. M. Sinaga, "Sistem Pakar Diagnosis Kondisi Kesehatan Mental Masyarakat Usia 18–23 Tahun Menggunakan Metode Certainty Factor dan Forward Chaining," Jurnal Teknologi Informasi, Komputer, dan Aplikasinya, vol. 5, no. 2, pp. 173–182, 2023, doi: https://doi.org/10.29303/jtika.v5i2.244.

[16] A. P. Widyassari, J. R. Carreon, and R. Wahyusari, "Systematic Review of Expert Systems for Detecting Mental Health Disorders in College Students," J. Tek. Inform. (JUTIF), vol. 5, no. 6, pp. 1737–1746, Dec. 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.6.4089.