Toward Trustworthy AI for Post-Stroke Suicidal Risk: Explainable Interaction Modeling and External Validation*

Flavia Costi[1], Emanuel Covaci[1], Darian Onchis[1], and Ovidiu Roman[2]

[1] Faculty of Informatics, Department of Computational Sciences and Artificial Intelligence, West University of Timisoara, Timisoara, Romania
{flavia.costi, emanuel.covaci, darian.onchis}@e-uvt.ro
[2] Clinical Hospital for Medical Rehabilitation, Baile Felix, Romania
romanovidiu22@yahoo.com

Abstract. Background: Suicidal risk after stroke is a serious clinical problem. Depression is a key predictor, but its effect may change when combined with anger expression and inflammation. It is also unclear whether the depression-related component of such a framework remains stable on external public data. **Methods:** We evaluated an explainable interaction-based framework on two datasets. The main analysis used a private cohort of 100 post-stroke patients, where suicidal risk was grouped into low, medium, and high levels using the Beck Hopelessness Scale. We modeled depression × anger_in, depression × anger_out, and depression × C-reactive protein (CRP) with regularized multinomial logistic regression.. We also analyzed a public NHANES cohort of 5,390 adults, where suicidal ideation was defined from PHQ-9 item 9 and depressive symptoms were represented by PHQ items. **Results:** In the private cohort, both models showed good discrimination. The most stable finding was the depression × anger_out interaction for the high-risk class ($\beta = 0.683$, 95% CI [0.429, 0.940], sign consistency 100%), while depression × anger_in was weaker but still positive. The depression × CRP interaction was suggestive but less robust. ALE stability further showed a highly consistent profile for outward anger ($\rho = 0.998$), in contrast to inward anger ($\rho = 0.103$). In NHANES, the depression-based model remained stable (AUC $= 0.909 \pm 0.019$, AP $= 0.306 \pm 0.053$), and the most stable predictors were DPQ060 and DPQ020, both with perfectly consistent ALE profiles ($\rho = 1.000$). **Conclusions:** Post-stroke suicidal risk appears to be shaped by interacting pathways, with depression as the central component and outward anger as the most consistent psychological amplifier in the high-risk group. The NHANES analysis supports the external stability of the depression-related pathway, whereas the interaction findings remain more specific to the post-stroke setting.

Keywords: post-stroke suicidal risk; depression; anger; C-reactive protein; explainable AI

* This research is partially supported by the project HRIA, Smart Growth, Digitization and Financial Instruments Program, 2021-2027, no. 334906.

1 Introduction

Stroke is one of the leading causes of disability and death worldwide, and its consequences extend beyond motor and cognitive impairment [1]. Emotional disturbances are common during recovery, and post-stroke depression is among the most frequent and clinically important of these complications [2]. Depression after stroke has been associated with poorer recovery, reduced quality of life, and increased suicidal risk [3, 4]. For this reason, suicidal risk should be considered a relevant component of post-stroke assessment rather than a secondary outcome.

Depression is a central predictor of suicidal risk, but it is unlikely to act alone. Psychological factors such as anger expression may alter how depressive symptoms translate into risk. Inward anger has been linked to self-blame and internalized distress, whereas outward anger has been associated with impulsivity, conflict, and emotional dysregulation [5, 6]. Biological factors may also contribute. Inflammation-related processes have been increasingly connected to post-stroke depression, and markers such as C-reactive protein, erythrocyte sedimentation rate, and fibrinogen may capture a state of higher vulnerability [7, 8]. Together, these observations suggest that suicidal risk after stroke may be better understood through interactions between depression, anger expression, and inflammation than through isolated predictors alone [9].

This creates both a modeling and an interpretation problem. Approaches based only on main effects may miss clinically meaningful interaction patterns, while more flexible predictive models often reduce interpretability and are harder to justify in clinical settings [11]. In suicide-risk research, good prediction is not sufficient on its own. Clinicians also need to know which variables drive risk and whether the identified patterns are interpretable and stable. Explainable AI is useful in this setting because it makes model behavior visible while allowing interaction-aware analysis [12]. In this study, we use Accumulated Local Effects and two-dimensional response maps to examine both main effects and joint patterns [13–15].

The aim of this study is to develop an explainable interaction-based framework for post-stroke suicidal-risk modeling and to evaluate which parts of this framework remain stable across datasets. We address this problem in two steps. First, in a private post-stroke cohort, we examine whether suicidal risk is shaped by the interactions depression $\times$ anger_in, depression $\times$ anger_out, and depression $\times$ CRP. Second, in a public NHANES cohort, we assess whether the depression-related component of the framework remains stable outside the post-stroke setting. Thus, the private dataset is used to study clinically specific interaction effects, whereas the public dataset is used to evaluate external stability of the transferable depressive pathway. The contribution of this work is an explainable and stability-aware framework that links prediction with interpretation while separating context-specific interaction signals from externally supported depressive risk patterns [16].

2 Materials and Methods

This study uses two datasets with different roles. The first is a private clinical dataset collected from hospitalized post-stroke patients, used to investigate suicidal risk in a disease-specific setting. The second is a public dataset derived from the National Health and Nutrition Examination Survey (NHANES), used to test whether the depression-related part of the proposed framework remains useful on external data. In this way, the private dataset supports the main clinical analysis, while the public dataset provides external support for the depression-related component of the framework.

The private dataset includes post-stroke patients evaluated at the *Clinical Hospital for Medical Rehabilitation*, Oradea, Romania. These are real patients assessed during the rehabilitation phase, typically 2–6 months after stroke onset. The private cohort included 100 post-stroke patients (64.9 $\pm$ 8.5 years; 60.0% women). Most cases were ischemic strokes (82.0%), while hemorrhagic strokes accounted for 18.0%. Suicidal-risk labels were unevenly distributed, with 71.0% low-risk, 26.0% medium-risk, and 3.0% high-risk cases. This cohort profile highlights both the clinical relevance of the dataset and the strong class imbalance at the highest-risk end. All data were anonymized before analysis according to institutional ethical standards and the Declaration of Helsinki. This dataset contains psychological, biological, clinical, and demographic variables. The main psychological variables include depression, anxiety, and anger expression. The biological and clinical variables include inflammatory markers, such as CRP, erythrocyte sedimentation rate, and fibrinogen, together with stroke-related and general clinical descriptors. Demographic variables such as age and sex were also available. In this private cohort, the outcome was an ordinal suicidal-risk proxy derived from the Beck Hopelessness Scale (BHS). Patients were assigned to low-, medium-, or high-risk groups according to increasing BHS severity, with higher scores indicating higher suicide-related risk. The aim of this dataset in the present study is to examine whether suicidal risk after stroke is shaped by interactions between depression and anger expression, and between depression and inflammation.

The public dataset is derived from NHANES [17]. After filtering for adult participants with valid PHQ responses and complete depressive symptom information, the final dataset included 5,342 subjects. Unlike the private cohort, this dataset is not stroke-specific and represents a broader adult population. Its role in this work is not to reproduce the clinical post-stroke setting, but to test whether the depression-related part of the proposed framework remains stable on external public data. In this dataset, the outcome variable was binary suicidal ideation, defined from PHQ-9 item 9. The predictor set included the PHQ depressive symptom items, together with demographic and lifestyle-related variables such as age, sex, race/ethnicity, education, marital status, smoking-related variables, alcohol-use variables, family income ratio, and body mass index. In additional experiments, blood-cell-derived inflammatory proxies were also considered.

Taken together, the two datasets support two complementary goals. The private post-stroke cohort is used to study clinically meaningful interaction effects, especially those involving depression, anger_in, anger_out, and CRP. The public NHANES cohort is used to test external portability of the depression-related component of the framework. The main focus of the paper is therefore not only prediction, but also the interpretation of how these variables shape suicidal-risk estimates across a clinical cohort and an external public cohort.

The two datasets were analyzed with different modeling strategies matched to their structure and role in the study. In both cases, the pipeline combined prediction with post-hoc explanation. For the private post-stroke cohort, we used regularized multinomial logistic regression. This choice was appropriate for three reasons. The sample was small. The outcome had three classes. The main objective was to estimate clinically meaningful interaction terms with a model that remained directly interpretable. We therefore explicitly constructed the interaction terms depression $\times$ anger_in, depression $\times$ anger_out, and depression $\times$ CRP. L2 regularization was used to reduce overfitting and improve coefficient stability [18]. Because the clinical question was centered on severe suicidal risk, coefficient stability was analyzed specifically for the high-risk class rather than averaged across classes. For the public NHANES cohort, the setting was different. The outcome was binary suicidal ideation, the sample was much larger, and the aim was external evaluation of the depression-related component rather than discovery of post-stroke-specific interaction mechanisms. In this case, we used HistGradientBoosting as the main model. This model was selected because it performed well on a larger tabular dataset with mixed symptom, demographic, and lifestyle predictors and could capture non-linear effects. Inflammatory proxies were examined as secondary additions, but the main public-data analysis remained centered on the depressive symptom profile [17].

Preprocessing was performed separately for each dataset and always inside the training data only. This avoided information leakage. In the private cohort, the analyzed variables had no missing values. Preprocessing therefore consisted of scaling continuous variables and encoding categorical variables. In NHANES, coded demographic variables were treated as categorical, while PHQ items were retained on their original ordinal scale. All preprocessing steps were recomputed at each split, fold, or bootstrap iteration.

Validation was also adapted to the role of each dataset. In the private cohort, we used repeated stratified 3-fold cross-validation. This avoided an unstable fixed test split and preserved class proportions as much as possible in a small multiclass sample. We then applied stratified bootstrap resampling to assess stability of the high-risk class coefficients. Resampling was performed within each class to ensure that all outcome levels, including the rare high-risk class, were represented in every bootstrap sample. Stability was summarized by the mean coefficient, standard deviation, percentile-based 95% confidence interval, and sign consistency across resamples. For the main psychological interaction, we also computed rank consistency, defined as the proportion of resamples in

which depression × anger_out exceeded the competing predictors and interaction terms.

In the public NHANES cohort, we used a stratified 80/20 train–test split. Model development was performed on the training data, and final evaluation was reported on the held-out test set. To quantify robustness, we estimated performance stability over repeated bootstrap splits. We also assessed permutation-importance stability across resamples and measured the consistency of ALE profiles for the dominant depressive predictors.

The main analytical workflows for the private and public datasets are illustrated in Figures 1 and 2. Performance was assessed with AUC-ROC, average precision, and Brier score. For the private multiclass task, AUC-ROC and average precision were computed in a one-vs-rest manner and macro-averaged across classes. For the public NHANES task, standard binary AUC-ROC, average precision, and Brier score were reported. These metrics were selected because they capture complementary aspects of model behavior: discrimination, precision under class imbalance, and probability calibration. Confidence intervals for the main stability analyses were obtained by bootstrap resampling [20].

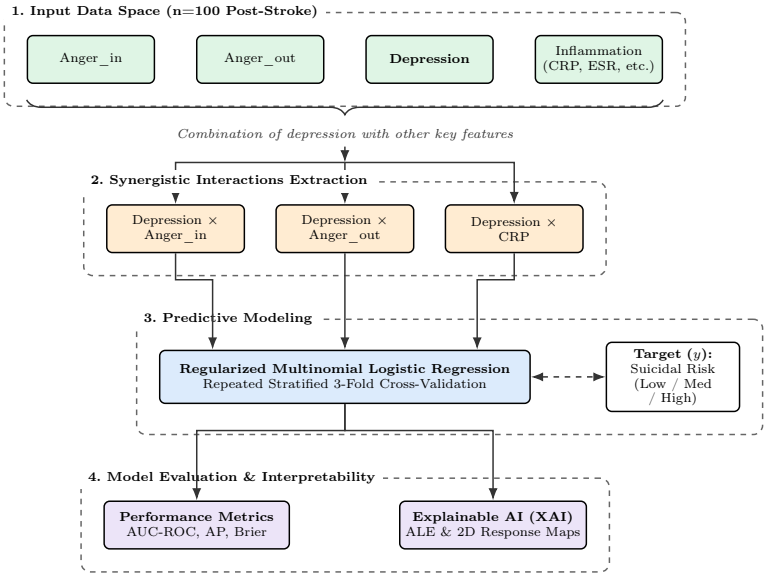


Fig. 1: Comprehensive analytical pipeline for the private post-stroke cohort. The layout uses conceptual grouping to illustrate how primary features are combined to extract synergistic interactions.

Explainability was treated as part of the main analysis. In the private cohort, interpretation focused on Accumulated Local Effects and two-dimensional response maps for the depression-related interactions [21, 22]. In the public NHANES

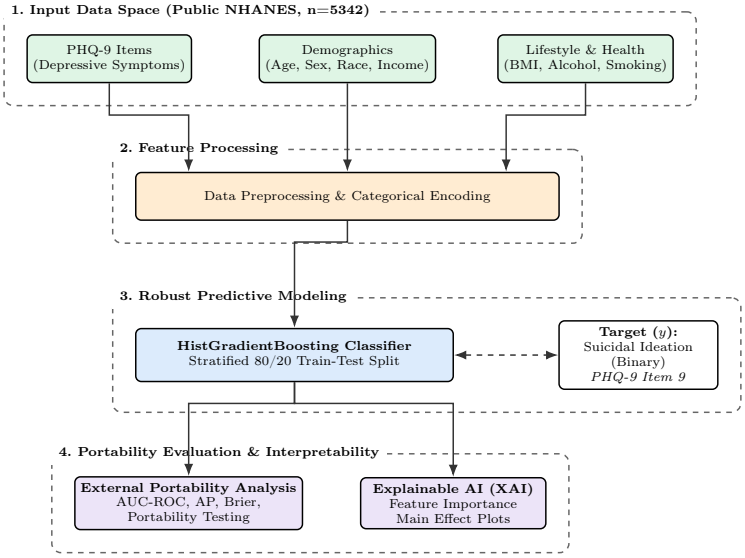


Fig. 2: Validation pipeline for the public NHANES cohort. The external portability workflow utilizes non-linear gradient boosting mapped directly to XAI validation.

cohort, interpretation focused on permutation importance and ALE main-effect profiles for the most influential PHQ predictors [23, 24]. To complement visual inspection, ALE profile stability was quantified with Spearman correlation between resampled curves. This allowed us to compare not only predictive performance, but also the stability of the explanatory patterns across the two datasets.

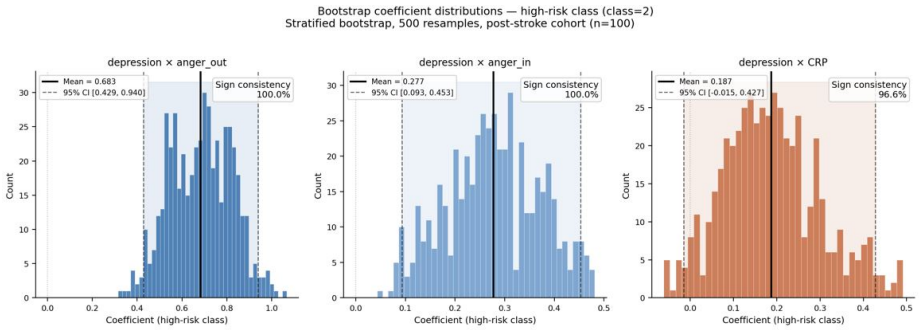
3 Results

The results are organized around the complementary roles of the two datasets. The private post-stroke cohort was used to examine clinically specific interaction patterns related to suicidal risk, whereas the public NHANES cohort was used to assess the external stability of the depression-related component of the framework. Accordingly, the Results section focuses on two questions: which interaction patterns are most relevant in the private cohort, and which components remain stable in the external public cohort.

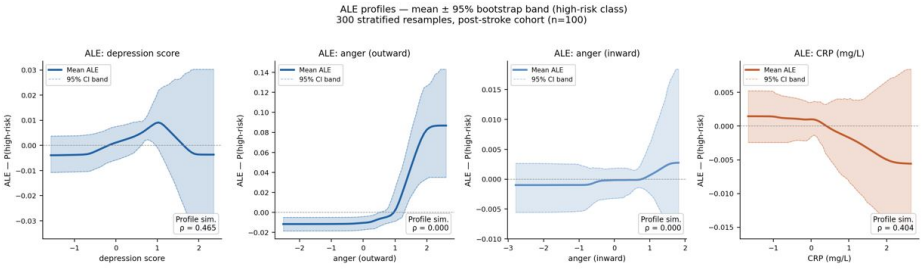
3.1 Private post-stroke cohort: interaction effects on suicidal risk

The private post-stroke cohort was used to examine whether suicidal risk is better explained by interaction effects than by isolated predictors alone. Two interaction pathways were analyzed. The first examined whether anger expression

modified the association between depression and suicidal risk. The second examined whether CRP altered the same depression-related pathway. In both settings, depression was treated as the reference clinical factor, while the interaction terms were used to test whether the probability of the high-risk class increased under specific psychological or biological conditions. Because the high-risk group was rare, interpretation focused not only on predictive performance, but also on the stability of class-specific coefficients and explanatory profiles. To provide a compact overview before the pathway-specific analyses, Figure 3a summarizes the bootstrap distributions of the main high-risk interaction coefficients, and Figure 3b shows the corresponding ALE profile stability across stratified bootstrap resamples. To provide a compact overview before the pathway-specific analyses, Figure 3a summarizes the bootstrap distributions of the main high-risk interaction coefficients, and Figure 3b shows the corresponding ALE profile stability across stratified bootstrap resamples.



(a) Bootstrap distributions of the high-risk class coefficients.



(b) ALE profile stability for the high-risk class.

Fig. 3: Stability analysis in the private post-stroke cohort. The top panel shows bootstrap distributions of the main interaction coefficients for the high-risk class. The bottom panel shows ALE profile stability across stratified bootstrap resamples. Depression × anger_out showed the strongest coefficient-level stability, and outward anger showed the most stable ALE profile.

To complement the importance analysis, and following the same stability-oriented perspective adopted in the private cohort, Figure 4 shows the ALE profile stability of the two dominant PHQ predictors across bootstrap resamples. This analysis was added to examine whether the depression-related explanatory patterns in NHANES remained reproducible, thereby providing stronger support for the external stability of the transferable depressive pathway.

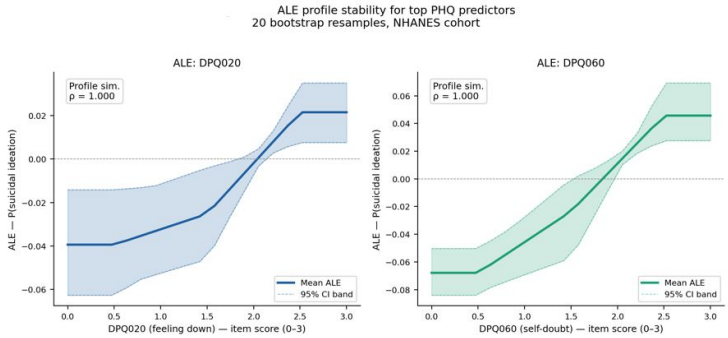


Fig. 4: ALE profile stability for the two dominant PHQ predictors in the public NHANES cohort. Both DPQ020 and DPQ060 showed perfectly consistent effect profiles across bootstrap resamples ($\rho = 1.000$), supporting the external stability of the depression-related pathway.

Psychological interaction: Depression × Anger The psychological model showed a clear difference between outward and inward anger. In the high-risk class, the strongest interaction was depression × anger_out, with a mean coefficient of 0.683, a 95% confidence interval of [0.429, 0.940], and sign consistency of 100% across stratified bootstrap resamples (Table 1).

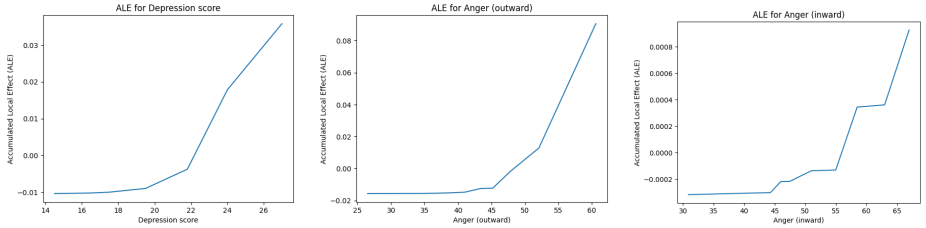
Predictor / Interaction	Mean β (high-risk)	SD	95% CI	Sign %
Depression	0.288	0.134	[0.081, 0.546]	100.0
Anger_out	0.510	0.062	[0.366, 0.615]	100.0
Anger_in	0.081	0.149	[-0.160, 0.359]	66.0
Depression × Anger_out	0.683	0.138	[0.429, 0.940]	100.0
Depression × Anger_in	0.277	0.094	[0.093, 0.453]	100.0

Table 1: High-risk class coefficient stability for the depression–anger model under stratified bootstrap resampling. The strongest and most stable term was depression × anger_out.

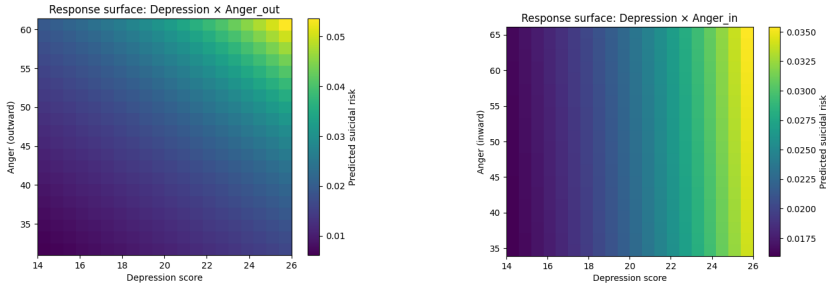
Measure	Value
Cross-validated AUC-ROC	0.924 ± 0.030
Depression × Anger_out sign consistency	100.0%
Depression × Anger_out > Depression × Anger_in	99.8%
Anger_out ALE similarity ρ	0.998
Anger_in ALE similarity ρ	0.103

Table 2: Summary of discrimination and stability for the depression–anger model. Outward anger showed both the strongest interaction with depression and the most stable explanatory profile.

The depression × anger_in interaction was also positive, but weaker (mean $\beta = 0.277$, 95% CI [0.093, 0.453]). The main effect of anger_out was positive and stable (mean $\beta = 0.510$, 95% CI [0.366, 0.615]), whereas anger_in alone was weak and unstable (mean $\beta = 0.081$, 95% CI [-0.160, 0.359], sign consistency 66.0%). Rank consistency analysis supported the same pattern: depression × anger_out exceeded depression × anger_in in 99.8% of resamples and exceeded the main effects of depression, anger_out, and anger_in in 100.0%, 95.0%, and 100.0% of resamples, respectively. These results identify outward anger as the dominant psychological amplifier of depressive vulnerability in the high-risk group. The corresponding model showed good discrimination under repeated stratified 3-fold cross-validation (AUC-ROC $= 0.924 \pm 0.030$; Table 2). Because the cohort was small and the high-risk class contained only three patients, this result was interpreted together with the class-specific bootstrap analysis rather than as a standalone performance estimate. The main result of the anger model was therefore not only predictive signal, but also the strong internal consistency of the depression × anger_out pathway.



(a) ALE - depression score. (b) Outward anger. (c) Inward anger.



(d) Depression × Anger_out. (e) Depression × Anger_in.

Fig. 5: Explainability results for the psychological interaction analysis in the private post-stroke cohort. The first row shows the ALE profiles for depression, outward anger, and inward anger. Outward anger shows the clearest and most stable effect. The second row shows the corresponding two-dimensional response surfaces. A marked high-risk ridge is visible for depression × anger_out, whereas depression × anger_in remains weaker and more additive.

The explainability analysis supported the same conclusion. The ALE profile for depression in Figure 5a showed an overall increase in the predicted probability of the high-risk class, although its resampled shape was only moderately stable ($\rho = 0.465$). In Figures 5b and 5c, outward anger showed a sharper and more stable pattern than inward anger. This difference was also visible in the two-dimensional response surfaces in Figures 5d and 5e. The depression × anger_out surface displayed a distinct high-risk ridge, whereas the depression × anger_in surface remained flatter and more additive. Taken together, the coefficient analysis, rank consistency, and ALE stability all pointed to outward anger as the most consistent psychological moderator in this cohort.

Biological interaction: Depression × CRP The biological model showed a weaker but still interpretable signal. In the high-risk class, depression remained positive and stable (mean $\beta = 0.505$, 95% CI [0.176, 0.893], sign consistency 100.0%; Table 3). The depression × CRP interaction was also positive on average (mean $\beta = 0.187$) and showed high sign consistency (96.6%), but its confidence interval crossed zero (95% CI [-0.015, 0.427]). In contrast, the main effect of CRP was small and unstable (mean $\beta = -0.052$, 95% CI [-0.252, 0.147], sign consistency 71.0%). This pattern suggests that CRP is not a robust standalone predictor of high suicidal risk, but may contribute when depressive burden is already elevated.

Predictor / Interaction	Mean β (high-risk)	SD	95% CI	Sign %
Depression	0.505	0.207	[0.176, 0.893]	100.0
CRP	-0.052	0.106	[-0.252, 0.147]	71.0
Depression × CRP	0.187	0.112	[-0.015, 0.427]	96.6

Table 3: High-risk class coefficient stability for the depression × CRP model under stratified bootstrap resampling. The interaction was positive on average, but less robust than the corresponding anger-related interaction.

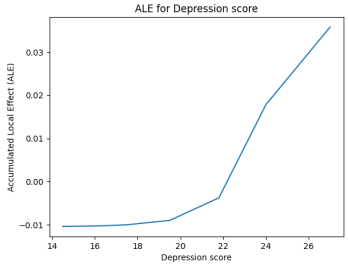
Measure	Value
Cross-validated AUC-ROC	0.917 ± 0.034
Depression × CRP sign consistency	96.6%
CRP main-effect sign consistency	71.0%
Depression ALE similarity ρ	0.465
CRP ALE similarity ρ	0.404

Table 4: Summary of discrimination and stability for the depression × CRP model. The interaction remained plausible, but weaker and less stable than the outward-anger pathway.

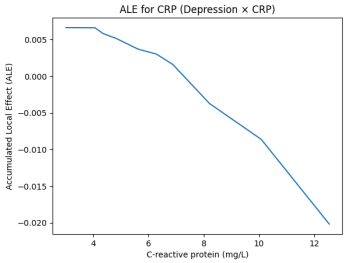
The depression–CRP model showed slightly lower discrimination than the anger-related model (AUC-ROC $= 0.917 \pm 0.034$; Table 4). The difference was modest at the predictive level, but the biological pathway was less stable at coefficient and ALE level. For this reason, the CRP-related findings should be interpreted as suggestive rather than definitive.

The explainability analysis was consistent with this interpretation. In Figure 6a, the ALE profile for depression again showed an overall positive trend. In Figure 6b, the ALE profile for CRP was weaker and less stable, indicating no clear main effect. However, the joint plots in Figures 6c and 6d suggested that higher CRP shifted the predicted probability of the high-risk class upward mainly at moderate-to-high depression values. Thus, the biological interaction

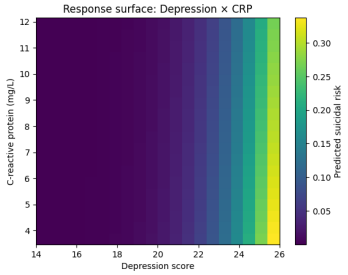
remained clinically plausible, but less robust than the psychological pathway driven by outward anger.



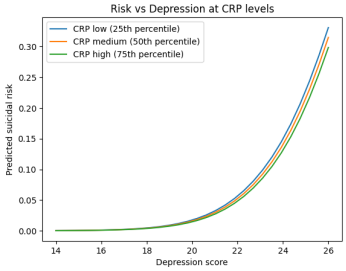
(a) ALE for depression.



(b) ALE for CRP.



(c) Joint surface for Depression × CRP.



(d) Risk curves by CRP level.

Fig. 6: Explainability results for the biological interaction analysis in the private post-stroke cohort. Depression retains a positive effect, whereas CRP shows a weaker and less stable main profile. The joint surface and conditional curves suggest that higher CRP shifts the predicted probability of the high-risk class upward mainly at moderate-to-high depression values.

3.2 Public NHANES cohort: external stability of the depression-related framework

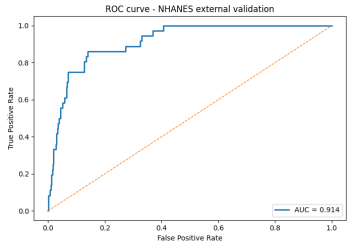
The NHANES analysis was used to test whether the depression-related component of the proposed framework remained stable on external public data. Unlike the private cohort, this analysis was not intended to reproduce the post-stroke setting or to validate the interaction structure observed there. Its role was to examine whether depressive symptom information retained predictive and explanatory value in a larger and more heterogeneous population.

The final public analysis was centered on the PHQ-based model. In the baseline evaluation, this model reached AUC-ROC = 0.845, AP = 0.222, and Brier = 0.0330. To quantify robustness, we then evaluated performance stability over 50 bootstrap splits. As shown in Table 5, the model remained stable, with mean

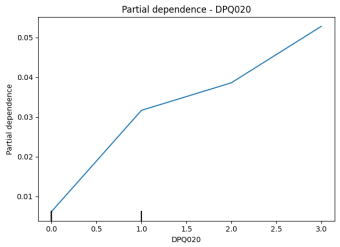
Metric	Mean	SD	95% CI
AUC-ROC	0.909	0.019	[0.867, 0.936]
Average Precision	0.306	0.053	[0.210, 0.390]
Brier Score	0.0291	0.0021	[0.0261, 0.0336]

Table 5: Performance stability of the public NHANES model over 50 bootstrap splits. The depression-based model showed stable discrimination and probability estimates under repeated resampling.

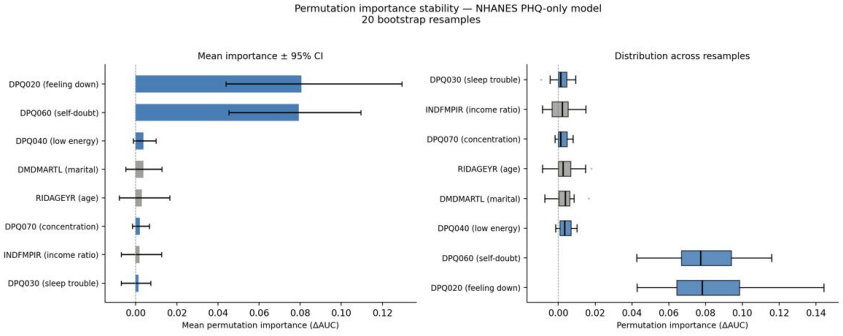
AUC-ROC of 0.909 ± 0.019, mean AP of 0.306 ± 0.053, and mean Brier score of 0.0291 ± 0.0021. These results indicate that the depression-based component retained good discrimination and stable probability estimates under repeated resampling.



(a) ROC curve for the NHANES model.



(b) Effect profile for DPQ020.



(c) Permutation importance stability across bootstrap resamples.

Fig. 7: External stability results on the public NHANES cohort. The model showed stable discrimination, while DPQ020 and DPQ060 emerged as the dominant and most stable depressive predictors.

The explainability analysis supported the same conclusion. Figure 7 summarizes the external stability of the public model from three complementary perspectives. The ROC curve confirms good discrimination, the effect profile of

DPQ020 shows a consistent monotonic increase in predicted suicidal-ideation risk, and the permutation-importance stability analysis identifies DPQ020 and DPQ060 as the two dominant predictors across bootstrap resamples. More specifically, both variables showed top-3 frequency of 100% and sign consistency of 100%, while their ALE profiles were also perfectly consistent ($\rho = 1.000$ for both). Additional subgroup analyses showed similarly strong performance across sex and age strata, further supporting the robustness of the public model. Taken together, these findings suggest that the transferable core of the framework is the depressive symptom pathway rather than the inflammation-related extensions.

4 Conclusions

This study proposed an explainable and stability-aware framework for suicidal-risk modeling, with a main focus on post-stroke patients and a complementary external evaluation on public data. In the private post-stroke cohort, the results suggest that suicidal risk may be better characterized by interacting pathways than by isolated predictors alone. Among the tested interactions, depression × anger_out showed the clearest and most stable signal for the high-risk class, whereas depression × anger_in remained weaker. The depression × CRP pathway appeared clinically plausible, but its evidence was less robust. Taken together, these findings suggest that depressive burden may be shaped by specific emotional and, to a lesser extent, biological modifiers in the post-stroke setting.

The public NHANES analysis served a different but complementary purpose. Although it did not reproduce the post-stroke setting, it provided external support for the depression-related component of the framework. The public model remained stable under repeated resampling, and the explanatory analysis consistently identified DPQ020 and DPQ060 as the dominant predictors. Their effect profiles were also highly reproducible, suggesting that the depressive symptom pathway may represent the most transferable part of the framework, whereas the interaction patterns observed in the private cohort appear to be more context-dependent.

A further contribution of this study is methodological. By combining predictive performance with explainability and stability analysis, the framework allowed a more cautious interpretation of both datasets. In the private cohort, this made it possible to distinguish between stronger and weaker interaction signals in the rare high-risk class. In the public cohort, it provided additional support that the main depressive predictors were not only important, but also reproducible across resamples. This improves the interpretability of the framework and strengthens the reliability of the reported findings.

The study also has important limitations. The private cohort was small, the high-risk group was extremely limited, and the biological findings were evaluated in a single clinical setting. In addition, the public dataset provided external support mainly for the depressive component of the framework rather than for the full post-stroke interaction structure. Future work should therefore include larger post-stroke cohorts, richer inflammatory profiling, and longitudinal follow-

up. Even with these limitations, the present results suggest that explainable and stability-aware modeling may offer a useful direction for more clinically grounded suicidal-risk assessment.

References

1. M. Mansoor, A. Ibrahim, A. Hamide, T. Tran, E. Candreva, and J. Baltaji, "Exercise-Induced Neuroplasticity: Adaptive Mechanisms and Preventive Potential in Neurodegenerative Disorders," *Physiologia*, vol. 5, no. 2, article 13, 2025. DOI: https://doi.org/10.3390/physiologia502001310.3390/physiologia5020013.
2. V. Brunetti, E. Rollo, I. Scala, J. Marotta, A. Callea, C. Imperatori, and G. Della Marca, "Sleep and Stroke-Related Delirium: A Systematic Review,"
3. M. Eriksson, E. L. Glader, B. Norrving, and K. Asplund, "Poststroke suicide attempts and completed suicides: a socioeconomic and nationwide perspective," *Neurology*, vol. 84, no. 17, pp. 1732–1738, Apr. 2015. DOI: https://doi.org/10.1212/WNL.0000000000000151410.1212/WNL.0000000000001514. PMID: 25832661; PMCID: PMC4424126.
4. C. W. Oei, E. Y. K. Ng, M. H. S. Ng, R.-S. Tan, Y. M. Chan, L. G. Chan, and U. R. Acharya, "Explainable Risk Prediction of Post-Stroke Adverse Mental Outcomes Using Machine Learning Techniques in a Population of 1780 Patients," *Sensors*, vol. 23, no. 18, article 7946, 2023. DOI: https://doi.org/10.3390/s2318794610.3390/s23187946.
   Clinical and Translational Neuroscience, vol. 7, no. 3, article 22, 2023. DOI: https://doi.org/10.3390/ctn703002210.3390/ctn7030022.
5. S. Selvaraj, S. Aggarwal, C. de Dios, J. M. De Figueiredo, A. Z. Sharrief, J. Beauchamp, and S. I. Savitz, "Predictors of suicidal ideation among acute stroke survivors," *Journal of Affective Disorders Reports*, vol. 10, p. 100410, 2022. DOI: https://doi.org/10.1016/j.jadr.2022.10041010.1016/j.jadr.2022.100410.
6. T. F. Hasan, H. Hasan, and R. E. Kelley, "Overview of Acute Ischemic Stroke Evaluation and Management," *Biomedicines*, vol. 9, no. 10, article 1486, 2021. DOI: https://doi.org/10.3390/biomedicines910148610.3390/biomedicines9101486. PMID: 34680603.
7. K. Kowalska, P. Pasinska, E. Klimiec-Moskal, J. Pera, A. Slowik, A. Klimkowicz-Mrowiec, and T. Dziedzic, "C-reactive protein and post-stroke depressive symptoms," *Scientific Reports*, vol. 10, no. 1, p. 1431, 2020. DOI: https://doi.org/10.1038/s41598-020-58478-610.1038/s41598-020-58478-6.
8. B. Goh and S. M. M. Bhaskar, "Evaluating Machine Learning Models for Stroke Prognosis and Prediction in Atrial Fibrillation Patients: A Comprehensive Meta-Analysis," *Diagnostics*, vol. 14, no. 21, article 2391, 2024. DOI: https://doi.org/10.3390/diagnostics1421239110.3390/diagnostics14212391. PMID: 39518359.
9. I.-B. Kim, J.-H. Lee, and S.-C. Park, "The Relationship between Stress, Inflammation, and Depression," *Biomedicines*, vol. 10, no. 8, article 1929, 2022. DOI: https://doi.org/10.3390/biomedicines1008192910.3390/biomedicines10081929.
10. B. Morgado, C. Silva, I. Agostinho, F. Brás, P. Amaro, L. Lusquinhos, M. R. Silva, C. Fonseca, N. Albacar-Riobóo, and L. G. de Pinho, "Psychotherapeutic Interventions for Depressive Symptoms in Community-Dwelling Older Adults: A Systematic Review with Meta-Analysis," *Healthcare*, vol. 12, no. 24, article 2551, 2024. DOI: https://doi.org/10.3390/healthcare1224255110.3390/healthcare12242551.

11. A. H. Miller and C. L. Raison, "The role of inflammation in depression: from evolutionary imperative to modern treatment target," *Nature Reviews Immunology*, vol. 16, no. 1, pp. 22–34, Jan. 2016. DOI: https://doi.org/10.1038/nri.2015.510.1038/nri.2015.5. PMID: 26711676; PMCID: PMC5542678.
12. Q. Abbas, W. Jeong, and S. W. Lee, "Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges," *Healthcare*, vol. 13, no. 17, article 2154, 2025. DOI: https://doi.org/10.3390/healthcare1317215410.3390/healthcare13172154.
13. E. Covaci, F. Costi, and D. M. Onchis, "ESM-VAE: Bias Reduction in EEG Models via Synthetic Data Generation with Variational Autoencoders," in *Intelligent Information and Database Systems*, N. T. Nguyen, T. Matsuo, F. L. Gaol, Y. Manolopoulos, H. Fujita, T.-P. Hong, and K. Wojtkiewicz, Eds., Singapore: Springer Nature Singapore, 2025, pp. 197–211. ISBN: 978-981-96-6008-7.
14. N. Kanwal, S. Javaid, and D. D. Dewi, "Stroke Prediction Using Machine Learning Algorithms," *Engineering Proceedings*, vol. 107, no. 1, article 32, 2025. DOI: https://doi.org/10.3390/engproc202510703210.3390/engproc2025107032.
15. S. Aksoy, P. Demircioglu, and I. Bogrekci, "Optimizing Stroke Classification with Pre-Trained Deep Learning Models," *Journal of Vascular Diseases*, vol. 3, no. 4, pp. 480–494, 2024. DOI: https://doi.org/10.3390/jvd304003610.3390/jvd3040036.
16. F. Costi, E. Covaci, and D. Onchis, "XplainLungSHAP: Enhancing Lung Cancer Surgery Decision Making with Feature Selection and Explainable AI Insights," *Surgeries*, vol. 6, no. 1, article 8, 2025. DOI: https://doi.org/10.3390/surgeries601000810.3390/surgeries6010008.
17. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Available online: `https://www.cdc.gov/nchs/nhanes/index.html` (accessed March 27, 2026).
18. Beck, A.T.; Steer, R.A.; Brown, G.K. *Manual for the Beck Depression Inventory-II*; Psychological Corporation: San Antonio, TX, USA, 1996.
19. Beck, A.T.; Weissman, A.; Lester, D.; Trexler, L. The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology* **1974**, *42*(6), 861–865.
20. Spielberger, C.D. *State-Trait Anger Expression Inventory: Professional Manual*; Psychological Assessment Resources: Odessa, FL, USA, 1988.
21. Spielberger, C.D.; Gorsuch, R.L.; Lushene, R.; Vagg, P.R.; Jacobs, G.A. *Manual for the State-Trait Anxiety Inventory*; Consulting Psychologists Press: Palo Alto, CA, USA, 1983.
22. A. G. Bottaccioli, M. Bologna, and F. Bottaccioli, "Psychic Life-Biological Molecule Bidirectional Relationship: Pathways, Mechanisms, and Consequences for Medical and Psychological Sciences—A Narrative Review," *International Journal of Molecular Sciences*, vol. 23, no. 7, article 3932, 2022. DOI: https://doi.org/10.3390/ijms2307393210.3390/ijms23073932. PMID: 35409300.
23. J.-S. Chou, D.-N. Truong, and C.-F. Tsai, "Solving Regression Problems with Intelligent Machine Learner for Engineering Informatics," *Mathematics*, vol. 9, no. 6, article 686, 2021. DOI: https://doi.org/10.3390/math906068610.3390/math9060686.
24. P. B. Le and Z. T. Nguyen, "ROC Curves, Loss Functions, and Distorted Probabilities in Binary Classification," *Mathematics*, vol. 10, no. 9, article 1410, 2022. DOI: https://doi.org/10.3390/math1009141010.3390/math10091410.