ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.281

Two-neural network based on the Goldberg scale for the classification of anxiety and depression

Jose Luis Ordoñez-Avila, Doctor en Dirección Empresarial[1,2], Elia Elizabeth Pineda Rivas, Master en Estadística aplicada a la investigación [1], and Vladimir Cruz Barahona, Master en Investigación Científica[1], and
[1]Universidad Evangélica de El Salvador, El Salvador, jose.ordonez@cvirtualuees.edu.sv, elia.pineda@uees.edu.sv, vladimir.cruz@uees.edu.sv
[2]Universidad Tecnológica Centroamericana (UNITEC), Honduras, jlordonez@unitec.edu

Abstract*– Anxiety and depression are highly prevalent and frequently comorbid mental disorders, suggesting the existence of shared cognitive and affective patterns. The Goldberg Anxiety and Depression Scale (GADS) is a widely used tool for their detection, supported by adequate validity and reliability indices in the Salvadoran population. However, the need to optimize screening processes justifies the search for methods that reduce the length of questionnaires without compromising their diagnostic capacity. The objective of this study was to evaluate whether the items of one GADS subscale are sufficient to predict the presence of symptoms in the opposite subscale using artificial neural networks. Two models were trained: one that used anxiety items to predict depressive symptoms and another that used depression items to predict anxiety. Both models showed solid performance, with F1-scores between 0.84 and 0.85 and AUC values above 0.82, demonstrating adequate discriminatory capacity. The results indicate that it is possible to significantly reduce the number of questions while maintaining accurate classification, which opens the door to the development of abbreviated versions based on deep learning for rapid mental health screening.

***Keywords--* Psychometric assessment, Predictive modeling, Symptom prediction.**

I. INTRODUCTION

Anxiety and depression are two of the most prevalent and clinically relevant mental disorders in the global population, characterized by significant alterations in cognition, emotional regulation, and social functioning. The World Health Organization defines mental disorders as clinically significant disturbances that cause distress, functional disability, or behavioral risk, emphasizing that their impact extends to multiple areas of daily life and represents a significant global disease burden [1]. This perspective allows us to understand the importance of having clear diagnostic criteria and valid and reliable assessment tools, especially in contexts where access to mental health services is limited.

In the specific case of anxiety, cognitive models—such as the one proposed by Beck and Clark—emphasize that it originates from exaggerated assessments of threat, attentional biases toward risk, and an underestimation of one's own coping resources [2]. From this framework, anxiety is fueled by selective processing of danger signals and inhibition of safety-related information, which contributes to maintaining the anxious state. Depression, meanwhile, has been described as a disorder characterized by persistent negative affect, anhedonia, and cognitive distortions associated with perceptions of loss, lack of control, and hopelessness, and is recognized as one of the main mental health problems in recent WHO reports [1]. The frequent comorbidity between anxiety and depression reinforces the hypothesis that both share underlying cognitive and affective mechanisms.

The current classification of anxiety disorders, as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), groups together various entities such as generalized anxiety disorder, panic disorder, specific phobias, social anxiety, agoraphobia, and other disorders induced by substances or other medical conditions [3]. Each has specific diagnostic criteria, but all share physiological, cognitive, and behavioral components that allow them to be considered within the same clinical spectrum. This clinical diversity highlights the need to use psychometric instruments with adequate validity and reliability properties, capable of capturing the different manifestations of anxiety and depression in specific populations.

In the Women Salvadoran context, the Goldberg Anxiety and Depression Scale (GADS) has established itself as one of the most widely used instruments for the simultaneous assessment of these symptoms. Similar studies address the issue of depression in women [4]. Local psychometric studies have shown that the adaptation of the GADS meets adequate construct validity criteria, with an explained variance of over 50% and acceptable confirmatory fit indices [5]. In methodological terms, various authors highlight the importance of mental health instruments having robust evidence of validity and reliability, so that they can objectively identify symptoms of anxiety and depression and provide comparable results across studies [6].

At the same time, growing interest in artificial intelligence in the clinical field has driven the development of models based on machine learning and deep learning to support the diagnosis and prediction of mental disorders. Recent literature shows successful applications of deep learning in detecting the risk of depression based on self-report questionnaires [7] and in the use of artificial intelligence techniques for various biomedical and psychological problems [8]. In this context, artificial neural networks offer a promising approach to optimizing screening processes [9-10], allowing us to explore whether it is possible to reduce the number of items in psychometric instruments without sacrificing their discriminatory power. In particular, the present study aims to evaluate whether the items of a GADS subscale are sufficient to predict the presence of symptoms in the opposite subscale using artificial neural networks, with the aim of moving towards abbreviated versions based on deep learning that facilitate rapid screening in mental health.

II. Methods

The study was conducted using a quantitative cross-sectional approach and aimed to evaluate the predictive capacity of artificial neural networks to identify symptoms of anxiety and depression using only the items from each subscale of the Goldberg Anxiety and Depression Scale (GADS). The design logic consisted of using the items corresponding to the anxiety subscale to train a model aimed at predicting the presence or absence of depressive symptoms, and then using the items from the depression subscale to train a second model aimed at predicting the presence or absence of anxiety. The dataset consist of 1367 women's from El Salvador.

The purpose of this strategy was to determine whether each set of items contains sufficient information to classify the other clinical condition, thereby reducing the total number of questions administered to participants. The data were collected using a digital application of the Goldberg Scale, an instrument consisting of 16 Likert-type items with four response options ranging from "Never" to "Always." Nine items correspond to the anxiety dimension and seven to the depression dimension. Participation was anonymous, voluntary, and the confidentiality of the information was guaranteed. To classify the participants, the cut-off points established for the scale were applied, considering the presence of anxiety or depression when the score reached or exceeded 18 points, and its absence when the scores were equal to or less than 17. These dichotomous variables were used as targets in the neural network models, while the individual items of each subscale were used as inputs. Initial processing included dividing the dataset into training and test subsets using a ratio of 80% and 20%, respectively, applying stratification to preserve class balance.

The model architecture was designed using a multilayer perceptron developed in TensorFlow/Keras. Each network included an input layer adjusted to the number of corresponding items (nine for the model that takes anxiety as input and seven for the model that takes depression), followed by four dense hidden layers with 16, 16, 8, and 8 neurons, respectively. The first hidden layer used the ReLU activation function, while the remaining layers used Swish activation to favor greater nonlinear representation capacity. In order to improve generalization and avoid overfitting, The output layer consisted of a single neuron with a sigmoid activation function, appropriate for binary classification tasks.

The models were trained using the Adam optimizer with a learning rate of 0.001, the Binary Crossentropy loss function, and the accuracy metric as the main performance indicator. Each model was trained for 100 epochs with a batch size of 32 samples, and the Early Stopping technique was included with a patience of 10 epochs to automatically stop training when the network stopped improving in validation.

Finally, performance evaluation was performed on the test set and included analysis of the confusion matrix, accuracy, precision, sensitivity (recall), and area under the ROC curve (AUC). In addition, the learning curves and the distribution of residual errors were examined, which allowed for the evaluation of training stability, generalization capacity, and consistency between model predictions and actual values. The results obtained made it possible to assess whether each subscale contains sufficient information to predict the other clinical condition, contributing to the possibility of reducing the total number of items necessary for classification without compromising diagnostic performance.

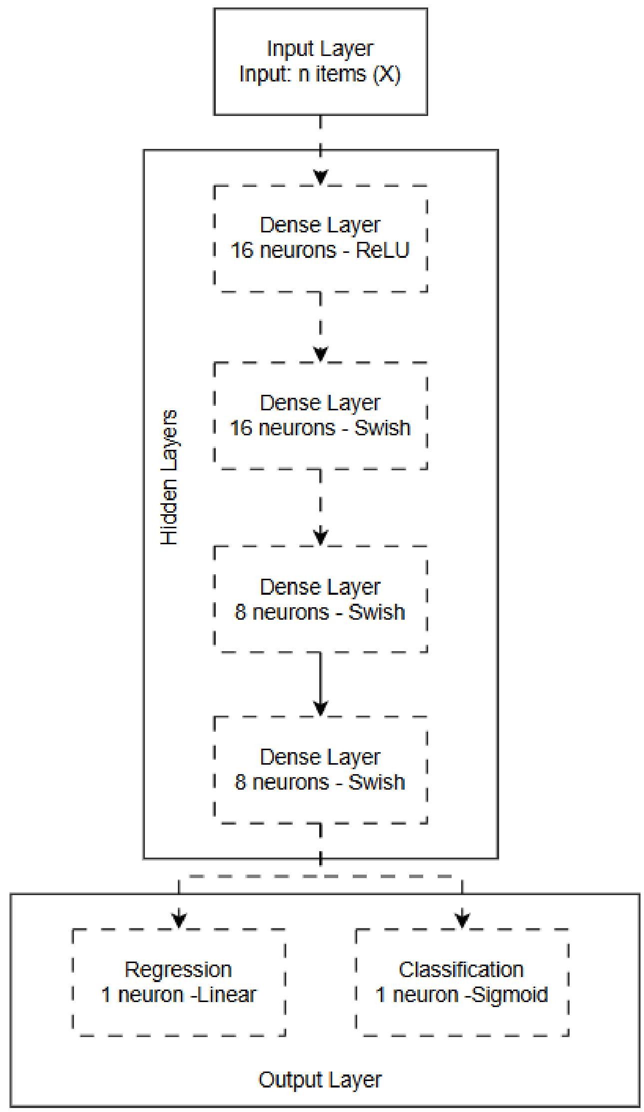

Fig. 1 Neural network architecture, with one input layer, four hidden layers with 16, 16, 8, and 8, and one output layer.

III. RESULTS

A. Input Anxiety- Output Depression

This section shows the results of the neural network that takes anxiety questions as input and depression classification as output. Figure 2 shows that total loss decreases rapidly during the early stages, falling from values above 400 to stabilize at around 10 from stage 20 onwards. This sharp drop indicates that the network learned useful patterns from the early stages of training. The proximity between the training curve and the validation curve shows that the model does not overfit, as both follow virtually parallel trajectories and converge toward a stable value. This confirms that the selected architecture has good generalization capabilities and that the optimization process worked properly.

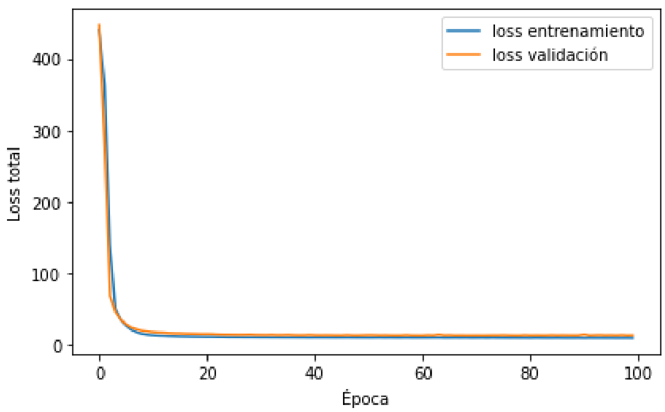

Fig. 2. Total loss evolution depression.

Figure 3 shows a significant reduction in absolute error from initial values close to 20 to a stable range between 2.5 and 3 points. The minimum distance between the training curve and the validation curve indicates that the model achieves a consistent fit in both sets, reinforcing the idea that there is no overfitting. A stable MAE around 2.5–3 is indicative of acceptable performance for a task of predicting depression scores based solely on anxiety items, suggesting that there is a correlation between the two subscales.

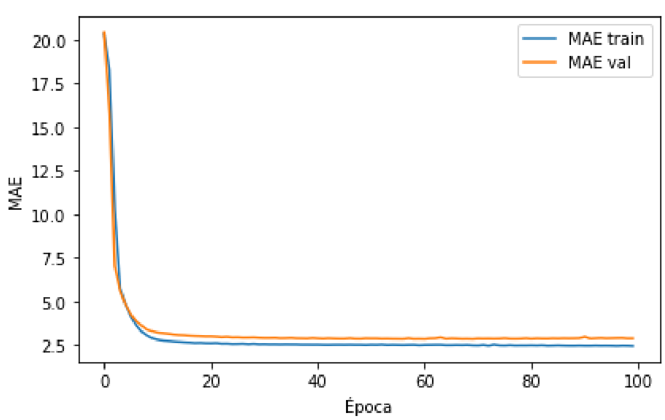

Fig. 3. Mean absolute error depression.

Classification accuracy shows (Fig. 4) a steady increase during the first 10 to 15 epochs, reaching values close to 0.85 for both training and validation. The validation curve remains within the same range as the training curve, and in some epochs even shows slightly higher performance, which is a desirable sign. An accuracy of around 85% indicates that the network is able to adequately differentiate between participants with and without symptoms of depression using only anxiety items, which is a solid result considering the reduced nature of the input vector.

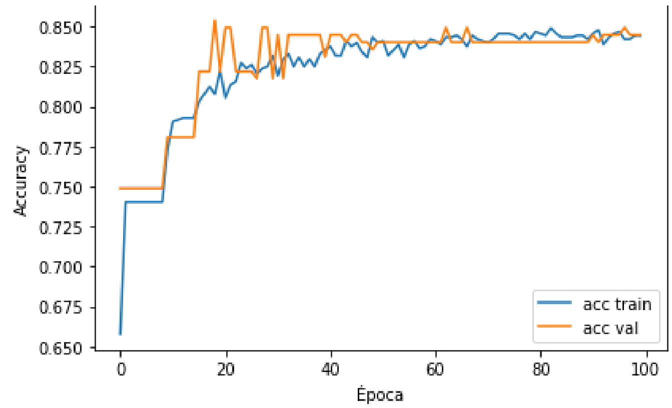

Fig. 4. Classification accuracy depression.

The ROC curve (Fig. 5) has an AUC=0.84, which is interpreted as good performance. An AUC of 0.84 indicates that the model has a solid ability to correctly separate participants with and without symptoms of depression. The curve remains above the random diagonal across almost the entire range of false positives, suggesting that the model captures real relationships between anxiety and depression and is not predicting arbitrarily.

Figure 6 shows an approximately linear relationship between actual and predicted scores, with acceptable dispersion around the trend line. Although there is some variability, especially at intermediate values, most points cluster around the diagonal, confirming that the network fits the continuous depression score estimate well. This supports the usefulness of the model as a hybrid estimator for simultaneous regression and classification.

The confusion matrix (Fig. 7) shows that the model correctly classified 175 participants with symptoms of depression and 39 without symptoms. However, it also shows that there were 38 false positives and 22 false negatives. Although the number of true positives is considerably high, the false positive rate implies that some people without symptoms were incorrectly classified as depressed. Nevertheless, the overall numbers suggest adequate predictive ability, with a good balance between sensitivity and accuracy.

The model's performance metrics for the classification of depressive symptoms were calculated using the confusion matrix. The accuracy obtained was 0.82, indicating that 82% of positive predictions corresponded to actual cases of depression. The sensitivity or recall reached a value of 0.89, showing that the model correctly identified 89% of people who actually had depressive symptoms, which is especially relevant in screening contexts where it is preferable to minimize false negatives. The F1-score, which represents the balance between accuracy and sensitivity, was 0.85, reflecting a robust and consistent performance of the classifier. Finally, the overall accuracy of the model was 0.78, implying that 78% of all predictions made, both positive and negative, were correct.

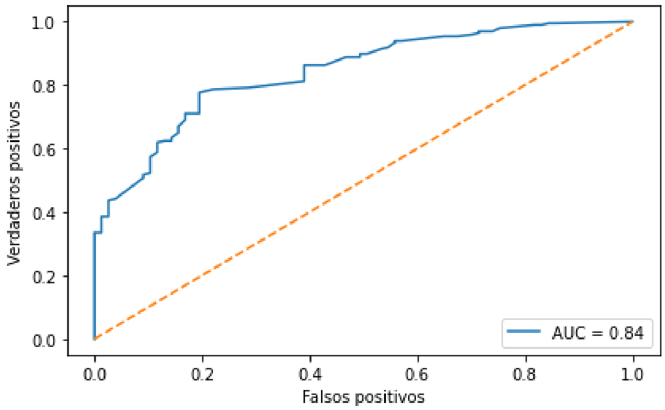

Fig. 5. Depression Receiver Operating Characteristic characteristics.

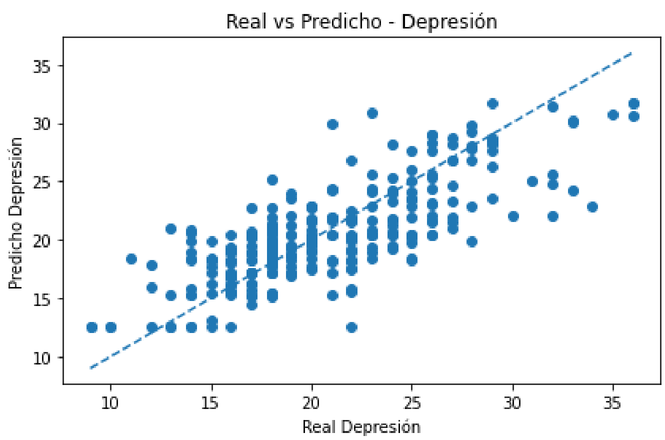

Fig. 6. Depression real vrs predicted dispersion analysis.

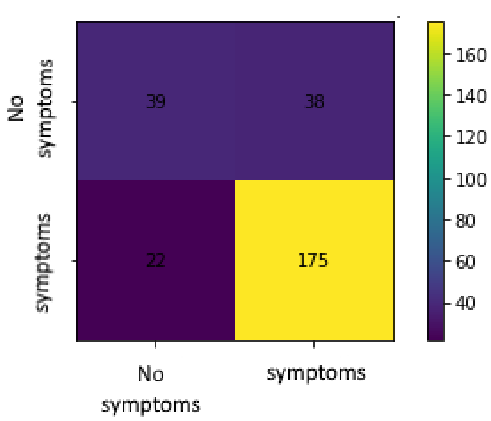

Fig. 7. Confusion matrix depression classification.

B. Input Depression - Output Anxiety

Figure 8 shows the total loss a rapid and pronounced decrease in the early stages, falling from values close to 200 to stabilizing below 10 around stage 20. This indicates that the network learned the relevant patterns efficiently from the start of training. The proximity between training loss and validation loss shows that the model remains stable and shows no signs of overfitting, as both curves remain practically superimposed throughout all epochs. This consistency suggests that the selected architecture has adequate generalization capacity when applied to new data.

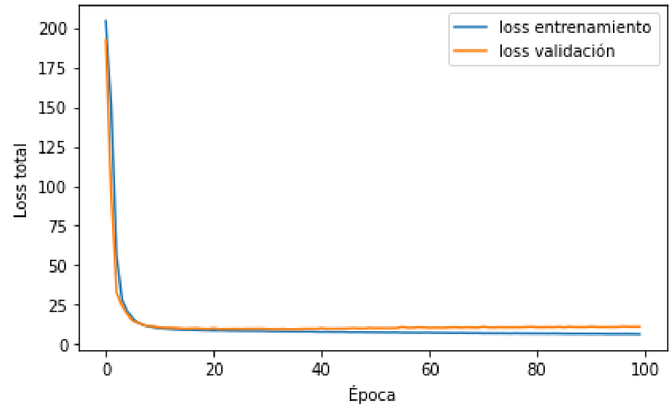

Fig. 8. Total loss evolution depression.

The MAE curve shows an accelerated reduction during the early stages, decreasing from initial values above 13 to stabilize between 2 and 3 points. Although the validation curve shows values slightly higher than those of training, it remains within a reasonable margin, indicating a good balance between learning and generalization. A final MAE close to 2.5–3 points is adequate considering that the model only uses depression items as input, demonstrating a consistent relationship between both dimensions of the GADS scale.

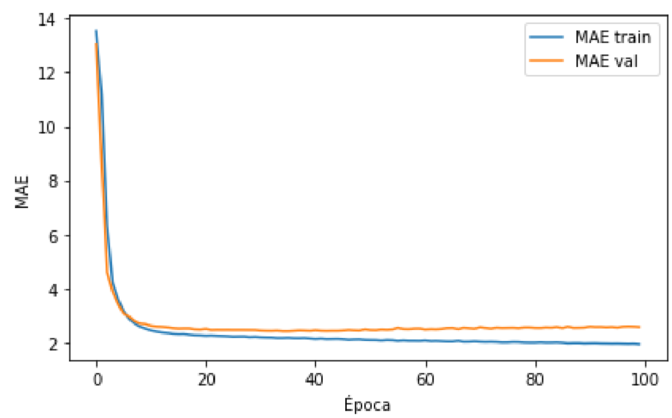

Fig. 9. Mean absolute error anxiety.

The accuracy of the model increases steadily to reach values between 0.83 and 0.86 in both training and validation. The stability observed in both curves, without abrupt drops or noticeable divergences, indicates that the model maintains healthy behavior and is not overfitted. These levels of accuracy reflect that the network is capable of adequately classifying the presence or absence of anxiety symptoms using only the depression items, which demonstrate predictive robustness and a strong interrelationship between both subscales.

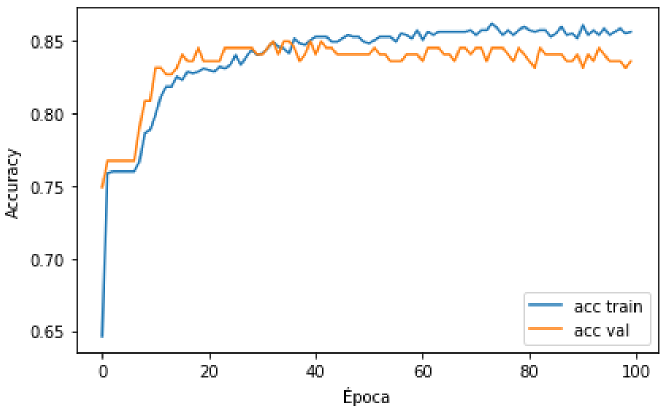

Fig. 10. Classification accuracy anxiety.

The ROC curve shows an area under the curve (AUC) of 0.82, which represents good performance. This value indicates that the model correctly distinguishes between positive and negative cases of anxiety in approximately 82% of possible comparisons. The curve remains above the random diagonal throughout the entire range, demonstrating that the model has strong discriminatory power and does not operate randomly. This result reinforces the validity of the approach of using one subscale to predict the other.

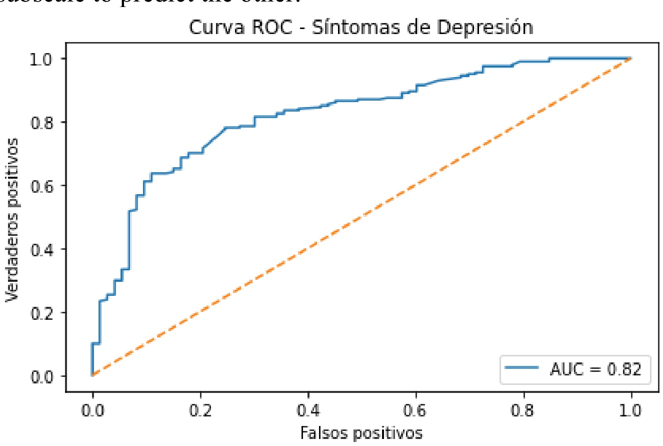

Fig. 11. Depression Receiver Operating Characteristic characteristics.

The figure showing actual versus predicted values shows a positive linear trend, although with greater dispersion than in the previous model. Most of the points are concentrated around the diagonal line, indicating that the network adequately captures the overall structure of the anxiety scores. The dispersion suggests that, although the model is useful, the prediction of continuous values has natural limitations when using only depression items, which is to be expected.

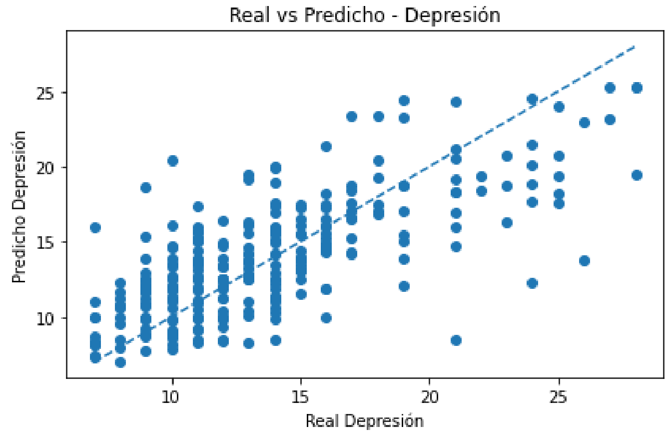

Fig. 12. Anxiety real vrs predicted dispersion analysis.

Based on the confusion matrix (Fig. 13), it was identified that the model correctly classified 32 people without anxiety symptoms (true negatives) and 176 people with anxiety symptoms (true positives), while producing 41 false positives and 25 false negatives. Based on these values, accuracy reached 0.81, indicating that 81% of positive predictions corresponded to actual cases. Sensitivity was 0.88, showing that the model correctly detected 88% of people who actually had anxiety, which is especially important in screening processes where minimizing false negatives is critical. The F1-score, which synthesizes the balance between precision and sensitivity, obtained a value of 0.84, reflecting a robust and balanced overall performance. Finally, the total accuracy reached 0.76, meaning that 76% of all predictions made were correct. Taken together, these metrics show that the model has an adequate ability to identify symptoms of anxiety using only depression items, with particularly strong performance in sensitivity, a desirable quality in early detection tools.

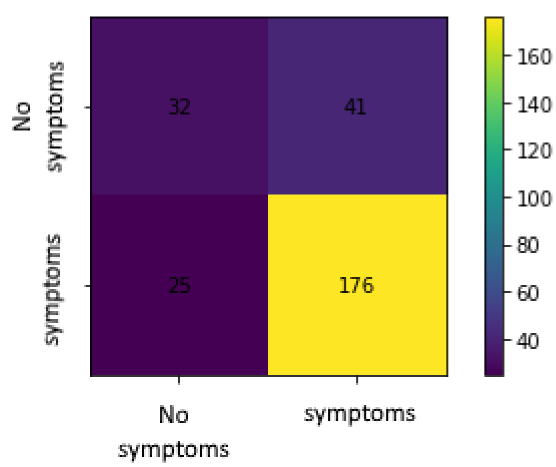

Fig. 13. Confusion matrix anxiety classification.

IV. Discussion

The results obtained in both neural networks show that the items in each subscale of the Goldberg Scale contain sufficient information to predict the other condition, suggesting a significant interrelationship between anxiety and depression symptoms. In the model that uses anxiety items to predict depression, the metrics show solid performance, with an F1-score of 0.85 and an AUC of 0.84, indicating correct discrimination between positive and negative cases. Similarly, the inverse model, which uses depression items to predict anxiety, obtained comparable indicators, with an F1-score of 0.84 and an AUC of 0.82, confirming that both dimensions share common patterns that allow accurate predictions even when the total number of questions is reduced. These metrics, together with the stability observed in the loss and accuracy curves, support the validity of the approach based on questionnaire reduction through deep learning.

Together, the two models demonstrate that the application of neural networks allows adequate diagnostic accuracy to be maintained even when a significant part of the original instrument is removed, opening the possibility of developing shorter and more efficient assessments without compromising detection capability. The high sensitivity observed in both models (0.88–0.89) is particularly relevant for mental health screening, where it is preferable to prioritize the detection of true cases to avoid missing individuals at clinical risk. However, the presence of false positives indicates that a reduced version should be used as a preliminary tool, subsequently complemented by a complete psychometric assessment. In summary, the results suggest that neural networks not only capture the internal structure of the GADS scale but also allow it to be optimized, offering a viable path toward more efficient screening instruments.

IV. Conclusion

Both models demonstrated that it is possible to predict symptoms of anxiety and depression using only items from the opposite subscale, maintaining solid performance with F1-scores above 0.84 and AUC between 0.82 and 0.84. This confirms a strong relationship between both dimensions and shows that neural networks can effectively reduce the number of questions without losing discriminatory power, supporting the creation of abbreviated versions of the Goldberg Scale for rapid mental health screening. These models are relevant to El Salvador due to the necessity of the depression and anxiety in women [11].

References

[1] World Health Organization, Depression and Other Common Mental Disorders: Global Health Estimates. Geneva, Switzerland: WHO, 2017.

[2] A. T. Beck and D. A. Clark, “An information processing model of anxiety: Implications for theory and treatment,” Behav. Res. Ther., vol. 35, no. 1, pp. 49–58, 1997.

[3] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th ed. Washington, DC, USA: APA, 2013.

[4] M. E. Lobos-Rivera and J. R. Gutiérrez-Quintanilla, “Adaptación psicométrica de la escala de ansiedad y depresión de Goldberg en una muestra salvadoreña,” Entorno, no. 70, pp. 87–98, 2020.

[5] O. A. Rodríguez-Reinoso, J. L. Ordóñez-Avila, A. G. Garay-Romero y R. M. Durón, “The contribution of women in engineering during the COVID-19 pandemic through indexed scientific publications,” en Proc. 3rd LACCEI Int. Multiconf. on Entrepreneurship, Innovation and Regional Development (LEIRD 2023), Virtual Edition, 2023

[6] E. E. Pineda Rivas, V. Cruz Barahona, and J. L. Ordoñez-Avila, “Symptoms of depression and anxiety in women according to occupation: A narrative bibliometric review,” in Proceedings of the Latin American Conference on Educational, Innovation and Research Development (LEIRD), 2025

[7] B. Choi, G. Shim, B. Jeong, and S. Jo, “Data-driven analysis using multiple self-report questionnaires to identify college students at high risk of depressive disorder,” Sci. Rep., vol. 10, no. 7867, 2020.

[8] L. Kouadio et al., “Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties,” Comput. Electron. Agric., vol. 155, pp. 324–338, 2018.

[9] K. A. Castro, J. A. Siwady, E. Castillo, A. Alonzo, M. Cardona and M. E. Perdomo, "Artificial Intelligence for All: Challenges and Harnessing Opportunities in AI Democratization," 2024 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), San Salvador, El Salvador, 2024, pp. 143-148, doi: 10.1109/ICMLANT63295.2024.00030.

[10] M. E. Perdomo, M. Cardona, D. M. Castro and W. M. Mejia, "Literature Review on Artificial Intelligence Implementation in the Honduran Agricultural Sector," 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Soyapango, El Salvador, 2022, pp. 1-5, doi: 10.1109/ICMLANT56191.2022.9996484.

[11] Pineda Rivas, E. E., Cruz Barahona, V., & Ordoñez-Avila, J. L. (2025). Symptoms of depression and anxiety in women according to occupation: A narrative bibliometric review. En Proceedings of the 5th LACCEI International Multiconference on Entrepreneurship, Innovation and Regional Development (LEIRD 2025). LACCEI. https://laccei.org/LEIRD2025-Cartagena/full-papers/Contribution_1124_final_a.pdf