

Modeling Social Anxiety, Attentional Bias, and Metacognitive Beliefs with Supervised Learning Algorithms

Josep. Bosch Borrás^{1*}, Zarley. Mett², Marlenne. Reynoso-Calderon³

¹ Department of Educational Psychology, School of Information Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, US

² Department of Psychology, Clark University, Worcester, Massachusetts, US

³ Department of Psychology, University of Toronto, Canada

* Corresponding author email address: boschborras-josep@gmail.com

Article Info

Article type:

Original Research

How to cite this article:

Bosch Borrás, J., Mett, Z., & Reynoso-Calderon, M. (2026). Modeling Social Anxiety, Attentional Bias, and Metacognitive Beliefs with Supervised Learning Algorithms. *Journal of Assessment and Research in Applied Counseling*, 8(1), 1-11.

<http://dx.doi.org/10.61838/kman.jarac.5181>



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Objective: The objective of this study was to utilize supervised machine learning algorithms to model and predict the severity of social anxiety based on the complex, non-linear interplay between multidimensional metacognitive beliefs and objective chronometric measures of attentional bias.

Methods and Materials: A cross-sectional, predictive study design was employed with 854 adult participants (mean age 34.52 years) recruited from the USA via online crowdsourcing. Social anxiety was assessed using the Liebowitz Social Anxiety Scale (LSAS), and metacognitive beliefs were measured via the Metacognitions Questionnaire-30 (MCQ-30). Attentional bias was quantified using a computerized visual dot-probe paradigm, measuring reaction times in milliseconds. Following preprocessing and k-nearest neighbors imputation for missing values (approximately 2%), the data were split into an 80%training and 20%testing set. Supervised learning algorithms, including Random Forest, Support Vector Machines, and Gradient Boosting Regressors, were trained using 10-fold cross-validation. Model evaluation relied on R^2 , Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), while feature importance was interpreted using SHapley Additive exPlanations (SHAP) values.

Findings: On the unseen testing set ($n = 171$), the Gradient Boosting Regressor emerged as the optimal model, demonstrating superior predictive performance ($R^2 = .68$, $RMSE = 13.65$, $MAE = 10.42$) compared to the Random Forest Regressor ($R^2 = .63$, $RMSE = 14.68$, $MAE = 11.55$) and the Support Vector Machine ($R^2 = .55$, $RMSE = 16.18$, $MAE = 12.89$). Feature importance analysis using SHAP values indicated that Negative Beliefs about uncontrollability and danger had the highest predictive influence (Mean | SHAP | = 6.84), closely followed by the Attentional Bias Score (Mean | SHAP | = 4.72) and the Need to Control thoughts (Mean | SHAP | = 3.55). Furthermore, SHAP dependence plots revealed a distinct non-linear interaction wherein the predictive

impact of attentional bias was exponentially magnified in the presence of high negative metacognitive beliefs.

Conclusion: Advanced machine learning algorithms successfully captured the complex architecture of social anxiety, demonstrating that clinical severity is predominantly driven by a synergistic, non-linear interaction between severe negative metacognitions and bottom-up visual hyper-vigilance toward threat.

Keywords: *Social Anxiety, Metacognitive Beliefs, Attentional Bias, Supervised Machine Learning*

1. Introduction

Social anxiety disorder (SAD) is widely recognized as one of the most prevalent and functionally debilitating psychiatric conditions globally. Characterized fundamentally by an intense, persistent fear of social or performance situations where the individual is exposed to unfamiliar people or possible scrutiny by others, SAD leads to pervasive avoidance behaviors. The functional impairment associated with this disorder is profound, often disrupting educational attainment, occupational functioning, and the basic ability to navigate daily societal demands. Furthermore, the psychosocial consequences are severe; individuals grappling with SAD frequently report a significantly diminished overall quality of life and experience pervasive, deeply rooted feelings of loneliness and isolation (Rajabi, 2018). The core psychopathology inherently impairs the development and maintenance of healthy social bonds, establishing a trajectory where specific cognitive distortions reliably predict subsequent interpersonal problems and relational dysfunctions (Strand et al., 2023). Because of the complex, multifaceted nature of its etiology and chronicity, achieving a comprehensive understanding of the psychological mechanisms that maintain social anxiety remains a critical objective in contemporary clinical psychology, requiring an intricate examination of overlapping cognitive, emotional, and behavioral vulnerabilities (Bagheri et al., 2024).

Historically, psychological conceptualizations of social anxiety heavily emphasized surface-level cognitive phenomena and historical developmental factors. For instance, extensive research has focused on the foundational role of maladaptive early schemas and their direct influence on symptom severity, often highlighting how these deep-seated core beliefs are mediated by various cognitive emotion regulation strategies (Ghaderi et al., 2022). While the specific content of a patient's thoughts is undeniably relevant to their clinical presentation, modern cognitive-behavioral paradigms have increasingly shifted their analytical focus toward the structural processes of cognition. It is now understood that variables representing how an

individual relates to their internal experiences—such as their dispositional mindfulness, their tendency toward experiential avoidance, and their specific fear of negative evaluation—are highly robust predictors of social anxiety severity (Rabat Mili & Karimi, 2018; Reshadat et al., 2019). This is particularly evident across different developmental stages; in younger populations and children, for example, the relative contribution of mere automatic negative thoughts to anxiety pathology is often vastly overshadowed by the destructive cognitive loops of repetitive negative thinking and higher-order metacognitive appraisals (Esbjörn et al., 2020). This distinct paradigm shift necessitates a much deeper investigation into the meta-level cognitive structures that actively govern and perpetuate anxious rumination and worry.

Central to this contemporary understanding is the theoretical framework of metacognition. Broadly defined as the psychological capacity for thinking about thinking, metacognition involves the higher-order monitoring, evaluation, and active regulation of one's own cognitive processes. According to the foundational Self-Regulatory Executive Function model, psychological distress is not directly caused by environmental triggers or initial negative thoughts, but rather by the persistent activation of a maladaptive cognitive pattern known as the Cognitive Attentional Syndrome (Aydm et al., 2025). This syndrome is characterized by an extended, inflexible pattern of self-focused attention, perseverative thinking through worry and rumination, and paradoxical coping behaviors that inadvertently backfire, such as attempting thought suppression. Crucially, the activation and persistence of the Cognitive Attentional Syndrome are strictly governed by underlying metacognitive beliefs (Z. Hashemi et al., 2022; Zohreh Hashemi et al., 2022). These include positive beliefs about the assumed utility of worrying as a coping mechanism, as well as highly distressing negative beliefs concerning the uncontrollability, inherent danger, and perceived social consequences of one's own thoughts. In many theoretical models, these metacognitive dimensions serve as a critical functional bridge, frequently mediating the complex psychological relationship between basic cognitive

emotion regulation strategies and higher-order states like mindfulness (Samia Shahani et al., 2020).

The detrimental impact of maladaptive metacognitive beliefs is not exclusively limited to social anxiety; rather, it is increasingly recognized within the literature as a foundational, transdiagnostic vulnerability factor present across a remarkably broad spectrum of human psychopathology. Metacognitive therapy, which directly targets these higher-order beliefs rather than modifying the specific content of daily worries, has demonstrated profound clinical effectiveness in ameliorating severe psychological symptoms. For instance, restructuring an individual's metacognitive architecture has been proven to significantly reduce pervasive worry and drastically improve the overall psychological well-being of patients suffering from generalized anxiety disorder (Abbasi et al., 2020). Similarly, in the domain of obsessive-compulsive disorder, interventions directly aimed at altering pathological metacognitive beliefs—such as the overestimation of the importance of intrusive thoughts—dramatically improve clinical outcomes and reduce compulsive rumination (Nateghi et al., 2019; Nematollahi et al., 2023). Furthermore, elevated dysfunctional metacognitions have been robustly linked to the manifestation and maintenance of intense delusional beliefs within severe obsessive-compulsive presentations (Qarali & saberi, 2020). The transdiagnostic reach of this conceptual framework extends deep into complex mood disorders, predicting the severity of cognitive deficits and emotional blunting in conditions like postpartum depression (Diop et al., 2022). Additionally, metacognitive deficits play a central, maintaining role in various behavioral addictions and impulse control issues, ranging from compulsive sexual behavior (Olivari et al., 2025) to the maladaptive tendency toward cyberspace addiction frequently observed even in highly functioning cohorts like gifted students (Asdolahzadeh et al., 2021).

Operating in parallel to the top-down metacognitive framework is the fundamental, bottom-up information processing mechanism of selective attention. Attentional bias refers to the systematic, highly unconscious tendency of the human brain to preferentially allocate visual and cognitive resources toward specific classes of stimuli in the surrounding environment, most notably stimuli that signal potential threat or interpersonal danger. Recent research rigorously exploring the underlying psychological predictors of attentional bias has demonstrated that this phenomenon is a complex construct extending far beyond the simplistic boundaries of basic trait anxiety, involving intricate

psychological interactions (Williams et al., 2024). In the specific context of social anxiety disorder, attentional bias typically manifests as a rapid visual hyper-vigilance toward socially threatening cues—such as faces expressing anger, disgust, or subtle rejection—immediately followed by a paradoxical, delayed inability to disengage visual attention from these threatening stimuli. This biased cognitive processing acts as a rigid perceptual filter, essentially guaranteeing that the socially anxious individual disproportionately registers negative social feedback while entirely missing neutral or positive social cues, thus constantly feeding highly biased, threatening data into their higher-order cognitive systems.

Much like metacognitive dysfunction, the phenomenon of threat-related attentional bias is a core, transdiagnostic processing mechanism deeply embedded within the underlying pathophysiology of numerous severe psychiatric conditions. The extreme clinical significance of attentional bias is heavily emphasized in trauma-related disorders, particularly Post-Traumatic Stress Disorder. In these specific populations, hyper-vigilance toward trauma-congruent stimuli actively maintains the disorder's chronicity, prompting the extensive development and meta-analytic evaluation of targeted clinical interventions such as Attentional Bias Modification and computerized attention control training paradigms (Zhang et al., 2024). The dysregulation of selective attention is equally prevalent and detrimental across the depressive spectrum. For example, highly sophisticated eye-tracking methodology has successfully isolated distinct, maladaptive patterns of attentional bias in expectant mothers suffering from antenatal depression, specifically highlighting how visual processing deficits directly contribute to the maintenance of low mood (Xu et al., 2023). Furthermore, precise chronometric cognitive assessments, such as the emotion-attention blink task, have definitively identified pervasive negative attentional biases in adolescent populations struggling with severe depressive disorders (Zhu et al., 2024). Across all these distinct conditions, the fundamental cognitive inability to flexibly direct visual attention away from negative stimuli serves as a primary, subconscious engine for continuous emotional dysregulation.

Despite the massive volume of empirical literature independently validating the clinical importance of metacognitive beliefs and the robust pathological effects of attentional bias, the complex, dynamic intersection between these two distinct cognitive domains remains a critical blind spot in current social anxiety research. Theoretically, these

systems must operate in a continuous, bidirectional psychological feedback loop. The Cognitive Attentional Syndrome, fundamentally driven by maladaptive metacognitive beliefs, inherently dictates the rigid allocation of finite cognitive resources. If an individual harbors severe negative metacognitions—such as the deeply held belief that their anxious thoughts are literally dangerous or represent a sign of impending psychological collapse—their entire neurocognitive system becomes primed for rapid threat detection. This profound metacognitive dread practically guarantees the execution of an attentional bias toward any external social cue that might validate their internal fears. Conversely, the continuous, biased registration of threatening faces via delayed attentional disengagement provides the exact empirical “evidence” required by the brain to reinforce and solidify the individual’s catastrophic metacognitive beliefs. Understanding precisely how these top-down metacognitive evaluations and bottom-up attentional processing metrics interact mathematically is absolutely vital for mapping the complete cognitive architecture of social anxiety.

Historically, the diverse methodological approaches deployed to investigate these intersecting psychological variables have relied almost entirely on traditional, parametric inferential statistics. Methodologies such as multiple linear regression analysis, hierarchical regression, and structural equation modeling conceptually assume that the complex relationships between independent psychological predictors and continuous clinical symptom severity can be adequately and accurately summarized by simple additive linear functions, mathematically denoted as $Y = \beta_0 + \sum_{i=1}^p \beta_i X_i + \epsilon$. While these conventional statistical frameworks have undeniably provided foundational, structural insights into clinical psychology, they are inherently limited by their rigid mathematical assumptions regarding perfect linearity, homoscedasticity, and the strict absence of severe multicollinearity among predictor variables. Human cognition, however, is rarely strictly linear. The intricate, functional relationship between a specific facet of metacognition, such as the urgent need to control thoughts, and the millisecond-level chronometric variations observed in computerized attentional bias tasks likely involves highly synergistic, non-linear mathematical interactions. Traditional additive models are generally incapable of natively capturing these multidimensional threshold effects, thereby potentially obscuring the true mathematical nature and complexity of the psychological data.

To effectively overcome the distinct mathematical limitations inherent in traditional psychological statistics, the strategic application of advanced supervised machine learning algorithms offers a highly powerful and sophisticated analytical paradigm. Unlike classical linear models, non-parametric machine learning techniques—such as Random Forest Regressors, Gradient Boosting ensembles, and Support Vector Machines—do not require a priori assumptions about the underlying statistical distribution of the clinical dataset. These predictive algorithms are specifically engineered to autonomously approximate highly complex, non-linear mathematical functions mapping a high-dimensional feature space X to a continuous target variable Y . By constructing mathematically complex decision boundaries and utilizing iterative ensemble learning techniques, these advanced algorithms can seamlessly identify and properly model intricate, deeply embedded interactive patterns among various metacognitive domains and precise attentional bias metrics. Furthermore, the integration of modern explainable artificial intelligence techniques, particularly the calculation of SHapley Additive exPlanations values, allows clinical researchers to mathematically deconstruct these dense algorithms. This provides a completely transparent, quantifiable measure of exactly how much mathematical weight every single cognitive predictor actively contributes to the final algorithmic prediction, offering an unprecedented, highly granular perspective on the intricate etiology of human anxiety.

Therefore, the precise aim of this study is to utilize an ensemble of advanced supervised machine learning algorithms to mathematically model and accurately predict the continuous severity of social anxiety based on the complex, non-linear interplay between multidimensional metacognitive beliefs and objective chronometric measures of attentional bias toward socially threatening stimuli.

2. Methods and Materials

2.1. Study Design and Participants

A cross-sectional, predictive study design was employed to investigate the complex relationships between social anxiety, attentional bias, and metacognitive beliefs using advanced supervised machine learning algorithms. The sample consisted of exactly 854 adult participants residing in the United States, recruited utilizing a stratified random sampling approach through prominent online crowdsourcing platforms dedicated to academic research. To ensure the

highest quality of data, strict inclusion criteria required participants to be between 18 and 65 years of age, demonstrate native-level fluency in the English language, and possess a verified current residency within the United States. Exclusion criteria involved any self-reported history of severe neurological disorders, psychotic episodes, or current uncorrected visual impairments that could potentially interfere with the visual demands of the computerized attentional bias tasks.

2.2. Measures

The rigorous assessment of the primary study variables was conducted utilizing a comprehensive battery of standardized self-report psychological questionnaires alongside a computerized cognitive behavioral task. Social anxiety symptomatology severity was quantitatively measured using the Liebowitz Social Anxiety Scale, a widely validated and highly reliable clinical instrument that carefully assesses both fear and avoidance behaviors across a multitude of distinct social interaction and performance situations measured on a continuous Likert-type scale. Metacognitive beliefs were systematically evaluated using the Metacognitions Questionnaire-30, an empirically supported tool that dynamically captures five core dimensions of metacognition, specifically including positive beliefs about worry, negative beliefs concerning the uncontrollability and danger of thoughts, cognitive confidence, beliefs regarding the urgent need to control thoughts, and cognitive self-consciousness. To accurately measure the phenomenon of attentional bias toward socially threatening stimuli, a precision-timed web-based iteration of the classic visual dot-probe paradigm was expertly programmed and administered. During this specific task, participants were momentarily presented with pairs of standardized facial stimuli consisting of one emotionally neutral face and one socially threatening or disgusted face, immediately followed by a visual target probe. Reaction times to probes replacing the threatening stimuli versus those replacing neutral stimuli were continuously recorded in milliseconds. Individual bias scores were mathematically calculated by subtracting the mean reaction time for congruent trials from the mean reaction time for incongruent trials, where all extremely rapid or delayed reaction times strictly falling outside the boundary of 200 to 1000 milliseconds were explicitly treated as statistical outliers and excluded from the final attentional bias score computations.

2.3. Data analysis

The compiled raw dataset underwent a rigorous and systematic preprocessing pipeline before the deliberate application of the supervised learning algorithms. Missing data values, which constituted approximately 2% of the total aggregate dataset, were effectively addressed and resolved using a *k*-nearest neighbors statistical imputation technique to maintain dataset integrity without introducing substantial bias. The core predictive modeling phase utilized a strategic ensemble of supervised machine learning algorithms, specifically deploying Random Forest, Support Vector Machines, and Gradient Boosting Regressors, to accurately model and predict the continuous severity of social anxiety based entirely on the inputted attentional bias indices and the multidimensional metacognitive belief scores. To prevent algorithmic overfitting and ensure robust generalizability, the entire dataset was randomly partitioned into a dedicated training set comprising exactly 80% of the data and a reserved testing set comprising the remaining 20%. Exhaustive hyperparameter tuning was meticulously performed exclusively on the training set using a 10-fold cross-validation methodology coupled with a randomized search mathematical strategy to empirically optimize the predictive performance of each individual algorithm. Final algorithmic model performance was strictly evaluated on the unseen testing set utilizing standard continuous evaluation metrics, principally including the coefficient of determination denoted mathematically as R^2 , the Root Mean Square Error, and the Mean Absolute Error. Furthermore, advanced post-hoc feature importance analyses, specifically utilizing SHapley Additive exPlanations values, were conducted to precisely quantify and interpret the individual and interactive statistical contributions of specific metacognitive belief facets and distinct attentional bias metrics to the overall algorithmic prediction of social anxiety severity.

3. Findings and Results

The initial phase of the data analysis focused on evaluating the fundamental characteristics of the sample and the descriptive statistics of the primary variables. The final sample consisted of $N = 854$ adult participants. Demographic analysis revealed that the mean age of the participants was $M = 34.52$ years with a standard deviation of $SD = 10.14$ years. The gender distribution was relatively balanced, comprising 54.2% females ($n = 463$), 43.6% males ($n = 372$), and 2.2% individuals identifying as

non-binary or other gender identities ($n = 19$). Regarding educational attainment, 42.5% held a bachelor's degree, 28.3% possessed a high school diploma or equivalent, 21.1% had completed a master's degree, and 8.1% held a doctoral or professional degree. Prior to training the machine learning models, zero-order Pearson correlation coefficients were computed to examine the bivariate linear relationships

between social anxiety severity, the five facets of metacognitive beliefs, and attentional bias scores. As anticipated, the Liebowitz Social Anxiety Scale total score demonstrated significant positive correlations with all measured predictor variables. The descriptive statistics and the complete intercorrelation matrix for all primary study variables are presented in Table 1.

Table 1

Descriptive Statistics and Zero-Order Correlations of Primary Study Variables

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7
1. Social Anxiety (LSAS)	58.34	24.12	–						
2. Attentional Bias (ms)	28.45	15.33	.41**	–					
3. MCQ: Positive Beliefs	14.22	4.51	.35**	.18*	–				
4. MCQ: Negative Beliefs	18.67	5.88	.62**	.31**	.44**	–			
5. MCQ: Cognitive Confidence	15.10	4.92	.48**	.22**	.29**	.51**	–		
6. MCQ: Need to Control	16.34	5.11	.55**	.28**	.38**	.60**	.45**	–	
7. MCQ: Cognitive Self-Consciousness	17.05	4.85	.43**	.25**	.33**	.48**	.39**	.52**	–

* $p < .05$. ** $p < .01$.

The correlational analysis revealed that among the metacognitive dimensions, Negative Beliefs concerning the uncontrollability and danger of thoughts exhibited the strongest bivariate relationship with social anxiety ($r = .62$, $p < .01$), followed closely by the Need to Control thoughts ($r = .55$, $p < .01$). Attentional bias toward socially threatening stimuli also demonstrated a robust positive correlation with social anxiety severity ($r = .41$, $p < .01$), indicating that individuals with higher social anxiety exhibited greater delayed disengagement from threatening faces. Importantly, the correlations among the predictor variables ranged from $r = .18$ to $r = .60$, indicating that while the variables are conceptually related, they do not violate the mathematical assumption of multicollinearity,

thus confirming their suitability for inclusion as distinct features within the supervised learning algorithms.

Following the preliminary analyses, the core predictive modeling phase was executed on the isolated 20% testing dataset ($n = 171$) to evaluate the generalization capabilities of the hyperparameter-tuned algorithms. The performance of the Random Forest Regressor, Support Vector Machine (SVM), and Gradient Boosting Regressor were systematically compared using three primary continuous evaluation metrics: the coefficient of determination (R^2), the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). The results of this rigorous comparative analysis are thoroughly detailed in Table 2.

Table 2

Performance Metrics of Supervised Learning Algorithms on the Unseen Testing Set

Predictive Model	R^2	RMSE	MAE
Gradient Boosting Regressor	.68	13.65	10.42
Random Forest Regressor	.63	14.68	11.55
Support Vector Machine	.55	16.18	12.89

As evidenced by the testing set performance metrics, the Gradient Boosting Regressor emerged as the mathematically superior model for predicting social anxiety severity. It successfully accounted for 68% of the variance in the Liebowitz Social Anxiety Scale scores ($R^2 = .68$). Furthermore, it achieved the lowest prediction error margins, with an RMSE of 13.65 and an MAE of 10.42, indicating

that its predictions deviated from the actual social anxiety scores by approximately 10 to 13 points on average. The Random Forest Regressor also demonstrated highly robust predictive capabilities, capturing 63% of the variance ($R^2 = .63$). In contrast, the Support Vector Machine yielded the lowest predictive accuracy among the ensemble, accounting for 55% of the variance ($R^2 = .55$) and exhibiting the

highest error rates. Consequently, the Gradient Boosting Regressor was unequivocally selected as the optimal algorithm for the subsequent feature importance extraction and post-hoc interpretability analyses.

To strictly quantify the specific contributions of attentional bias and the distinct metacognitive beliefs to the algorithmic prediction of social anxiety, SHapley Additive exPlanations (SHAP) values were computed based on the optimized Gradient Boosting Regressor model. SHAP

values provide a mathematically rigorous method for interpreting complex machine learning models by assigning an importance value to each feature for every individual prediction. The mean absolute SHAP values, representing the overall global importance of each feature in driving the model’s predictions, were extracted and ranked. Table 3 presents the precise hierarchical importance of these predictive features.

Table 3

Feature Importance Scores (Mean Absolute SHAP Values) for the Optimal Gradient Boosting Regressor

Predictor Variable	Mean SHAP Value	Importance Rank
MCQ: Negative Beliefs	6.84	1
Attentional Bias Score	4.72	2
MCQ: Need to Control	3.55	3
MCQ: Cognitive Confidence	2.10	4
MCQ: Cognitive Self-Consciousness	1.45	5
MCQ: Positive Beliefs	0.88	6

The SHAP analysis definitively established that Negative Beliefs regarding the uncontrollability and danger of thoughts constituted the paramount driving feature within the model, yielding the highest mean absolute SHAP value of 6.84. This signifies that variations in this specific metacognitive domain exerted the most profound mathematical impact on the model’s final prediction of social anxiety severity. Crucially, the computerized Attentional Bias score emerged as the second most vital predictor (Mean |SHAP| = 4.72), demonstrating that objective cognitive processing metrics significantly enhance predictive accuracy beyond self-reported metacognitive traits alone. The metacognitive Need to Control thoughts (Mean |SHAP| = 3.55) and Cognitive Confidence (Mean |SHAP| = 2.10) also provided substantial predictive utility. Conversely, Positive Beliefs about worry demonstrated the least predictive influence on the model’s output (Mean |SHAP| = 0.88). Further inspection of the SHAP dependence interactions, though not tabulated, revealed a critical non-linear interaction: the algorithmic penalty (increased predicted anxiety) associated with high attentional bias was exponentially magnified in participants who concurrently exhibited SHAP values corresponding to the upper quartile of Negative Metacognitive Beliefs. This indicates that the synergistic presence of rigid attention toward social threats and high negative metacognition produces a compounding algorithmic effect on predicted social anxiety severity.

4. Discussion

The primary objective of this current investigation was to mathematically model and meticulously predict the continuous severity of social anxiety by utilizing an advanced ensemble of supervised machine learning algorithms, specifically focusing on the complex, non-linear interplay between multidimensional metacognitive beliefs and objective, chronometric measures of attentional bias toward socially threatening stimuli. The algorithmic performance evaluation unequivocally demonstrated that the Gradient Boosting Regressor was the mathematically superior model, successfully accounting for 68% of the variance in social anxiety severity, denoted mathematically as $R^2 = .68$. This robust predictive capability significantly surpasses the typical variance explained by traditional linear regression models frequently utilized in psychiatric research. The extraction of SHapley Additive exPlanations values provided a highly transparent, quantifiable hierarchy of feature importance, revealing that negative metacognitive beliefs concerning the uncontrollability and inherent danger of thoughts, alongside the precise computerized attentional bias scores, were the paramount drivers of the algorithmic predictions. These findings provide substantial empirical support for contemporary cognitive-behavioral paradigms that emphasize the structural processes of cognition over the mere content of automatic thoughts, illustrating that the architecture of social anxiety is maintained by a highly

synergistic relationship between top-down metacognitive evaluations and bottom-up visual processing mechanisms.

The emergence of negative metacognitive beliefs as the single most critical predictive feature within the machine learning model profoundly aligns with and extends existing psychological literature. Extensive previous research has consistently highlighted the foundational role of maladaptive metacognitions in predicting the severity and maintenance of social anxiety (Reshadat et al., 2019). Unlike simple negative automatic thoughts, which often play a secondary role in advanced chronicity (Esbjörn et al., 2020), deep-seated negative beliefs regarding the danger of one's own cognitive processes actively construct a psychological environment characterized by pervasive threat and internal instability. This maladaptive cognitive framework directly predicts severe interpersonal problems and relational dysfunctions that are hallmark features of social anxiety disorder (Strand et al., 2023). Furthermore, the predictive dominance of these specific metacognitions aligns seamlessly with the theoretical framework of the Cognitive Attentional Syndrome, which posits that psychological distress is perpetuated by inflexible, self-focused attention and perseverative thinking driven by these exact underlying beliefs (Aydın et al., 2025). The structural modeling of social anxiety has repeatedly demonstrated that metacognitive beliefs often act as a critical functional bridge, mediating the relationship between early maladaptive schemas and subsequent emotional dysregulation (Ghaderi et al., 2022), and playing a pivotal role in the relationship between cognitive emotion regulation and higher-order states like mindfulness (Rabat Mili & Karimi, 2018; Samia Shahani et al., 2020).

Importantly, the transdiagnostic nature of these destructive negative metacognitions is highly supported by the broader clinical literature, reflecting their importance beyond just social anxiety. Interventions directly targeting these higher-order beliefs, such as metacognitive therapy, have demonstrated profound clinical effectiveness in ameliorating severe psychological symptoms across various disorders, significantly improving the quality of life and reducing feelings of loneliness in socially anxious patients (Rajabi, 2018). Similar therapeutic approaches have proven highly successful in reducing pervasive worry and improving psychological well-being in generalized anxiety disorder (Abbasi et al., 2020), as well as drastically improving clinical outcomes and reducing compulsive rumination in obsessive-compulsive disorder (Nateghi et al., 2019; Nematollahi et al., 2023). Dysfunctional

metacognitions are also robustly linked to intense delusional beliefs (Qarali & saberi, 2020), the severity of cognitive deficits in postpartum depression (Diop et al., 2022), and maladaptive behavioral tendencies ranging from cyberspace addiction in gifted populations (Asdolahzadeh et al., 2021) to compulsive sexual behaviors (Olivari et al., 2025). The fact that negative metacognitive beliefs heavily dictated the machine learning model's output in the present study solidifies their position as a universal vulnerability factor that drastically compounds social anxiety symptomatology. The metacognitive need to control thoughts also emerged as a highly significant predictor, further validating interventions that focus on executive function training and mitigating the Cognitive Attentional Syndrome in socially anxious individuals (Bagheri et al., 2024; Z. Hashemi et al., 2022; Zohreh Hashemi et al., 2022).

The second major finding of this study relates to the powerful predictive utility of the objective attentional bias score, which emerged as the second most vital feature in the Gradient Boosting Regressor. This mathematically confirms that cognitive processing metrics significantly enhance predictive accuracy beyond what can be achieved with self-reported metacognitive traits alone. The robust influence of attentional bias on the model's predictions strongly supports recent literature arguing that biased visual allocation is a complex construct extending far beyond basic trait anxiety (Williams et al., 2024). The inability to flexibly disengage from socially threatening facial stimuli serves as a continuous, subconscious engine for emotional dysregulation. This mechanism is not isolated to social anxiety; maladaptive patterns of selective attention and hyper-vigilance are core pathological features deeply embedded within trauma-related conditions like PTSD, where attentional control training is highly utilized (Zhang et al., 2024). Similarly, distinct patterns of visual processing deficits directly contribute to the maintenance of low mood in expectant mothers with antenatal depression (Xu et al., 2023) and adolescent populations struggling with severe depressive disorders (Zhu et al., 2024). The SHAP dependence analysis in the current study revealed a critical non-linear interaction: the algorithmic penalty for high attentional bias was exponentially magnified in participants who concurrently exhibited severe negative metacognitive beliefs. This suggests that the rigid allocation of visual attention toward social threats acts as a rapid, bottom-up mechanism that continuously feeds threatening empirical data to the brain, which then serves to intensely validate and solidify the individual's catastrophic top-down

metacognitive evaluations. This complex, compounding mathematical relationship highlights the necessity of utilizing advanced non-parametric modeling techniques in modern psychiatric research.

5. Conclusion

In conclusion, this study successfully leveraged advanced supervised machine learning algorithms to elucidate the complex, multidimensional architecture of social anxiety. By strictly evaluating the mathematical contributions of various psychological predictors, the investigation definitively established that the cognitive pathology of social anxiety is primarily governed by a highly synergistic interaction between maladaptive negative metacognitive beliefs and delayed visual disengagement from threatening social stimuli. The superior predictive performance of the Gradient Boosting Regressor, quantified by an exceptionally high coefficient of determination, underscores the profound mathematical limitations of traditional linear modeling when attempting to map intricate human cognition. Ultimately, these findings validate contemporary psychological theories that prioritize the structural mechanisms of thought and attention, proving that high-level metacognitive dread and low-level visual hyper-vigilance compound non-linearly to drive severe clinical symptomatology.

6. Limitations & Suggestions

Despite the robust methodological framework and utilization of advanced predictive modeling, several distinct limitations must be carefully considered when interpreting these findings. Foremost, the cross-sectional nature of the study design inherently precludes the establishment of definitive temporal causality between the measured cognitive variables and the continuous severity of social anxiety. While the machine learning algorithms successfully generated a highly accurate mathematical prediction based on concurrent data, the exact developmental trajectory of these intersecting psychological vulnerabilities remains entirely inferred. Furthermore, the reliance on a web-based crowdsourcing platform for data collection introduces potential sampling biases, and the strict inclusion of only adult residents within the United States significantly limits the broad cultural and developmental generalizability of the algorithms. Finally, while attention checks were utilized, self-reported metrics are always susceptible to subjective response biases.

To address these limitations and further advance the field, future research should urgently prioritize the execution of extensive longitudinal study designs that track the developmental fluctuations of metacognitive beliefs and attentional bias over extended temporal periods. The integration of highly sophisticated physiological data, such as high-frequency eye-tracking technology and electroencephalography, alongside machine learning algorithms would provide an incredibly granular, multi-modal assessment of the precise chronometric interactions occurring between visual perception and cognitive evaluation. Additionally, future investigations must seek to replicate and rigorously validate these specific predictive algorithms across diverse clinical cohorts, including child and adolescent populations, as well as individuals representing vastly different socio-cultural backgrounds. Testing these non-linear models in actively diagnosed clinical samples rather than general community populations will be critical for determining their true diagnostic utility.

The profound predictive significance of both metacognitive beliefs and attentional bias yields critical suggestions for modern psychological practice and clinical intervention. Mental health professionals treating social anxiety should transition away from solely targeting the specific content of negative automatic thoughts, and instead rigorously assess and directly restructure the patient's higher-order negative metacognitions regarding the danger and uncontrollability of their own cognition. The evidence of a compounding, non-linear interaction between cognitive domains strongly suggests that isolated therapeutic approaches may be inherently insufficient for severe presentations. Therefore, clinical practice should move toward highly integrative treatment protocols that simultaneously combine established metacognitive therapy techniques with rigorous, computerized attentional bias modification training. By concurrently dismantling the top-down catastrophic beliefs and correcting the bottom-up visual threat processing deficits, clinicians can effectively disrupt the complete cognitive feedback loop that systematically maintains functional impairment in socially anxious individuals.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

Authors' Contributions

All authors equally contributed in this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

References

- Abbasi, A., Aghaei Jeshvaghani, A., & Ebrahimi moghasam, H. (2020). Effectiveness of metacognition therapy on the anxiety and psychological wellbeing of the patients with generalized anxiety. *medical journal of mashhad university of medical sciences*, 63(5), 2745-2756. <https://doi.org/10.22038/mjms.2021.17348>
- Asdolahzadeh, P., Sadeghi, J., & Abbasi Asfajir, A. A. (2021). Modeling the Structural Equations of Mode Metacognition with a Tendency to Cyberspace Mediated by Self-efficacy in Gifted Students [Research Article]. *Iranian Journal of Educational Sociology*, 4(2), 14-23. <https://doi.org/10.52547/ijes.4.2.14>
- Aydın, S., Batmaz, S., & Aslan, E. A. (2025). Comparison of Cognitive Attentional Syndrome and Generic and Psychosis-Specific Metacognitive Beliefs in Remitted Patients with Schizophrenia, Schizoaffective Disorder, and Bipolar Disorder Type I with Psychotic Features with Healthy Controls. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 43(2), 1-17. <https://doi.org/10.1007/s10942-025-00586-1>
- Bagheri, A., Goodarzi, K., Rouzbehani, M., & Kakabraei, K. (2024). Comparing the effectiveness of metacognitive therapy, solution-focused therapy, and endurance exercises on self-efficacy beliefs in female students with social anxiety disorder. *Journal of Psychological Science*, 23(137), 133-153. <https://psychologicalscience.ir/article-1-2420-en.html>
- Diop, S., Turmes, L., Specht, C., Seehagen, S., Juckel, G., & Mavrogiorgou, P. (2022). Capacities for meta-cognition, social cognition, and alexithymia in postpartum depression. *Psychiatry research*, 309, 114430. <https://doi.org/10.1016/j.psychres.2022.114430>
- Esbjörn, B. H., Falch, A., Walczak, M., Normann, N., & Breinholst, S. (2020). Social Anxiety Disorder in Children: Investigating the Relative Contribution of Automatic Thoughts, Repetitive Negative Thinking and Metacognitions. *Behavioural and Cognitive Psychotherapy*, 49(2), 159-171. <https://doi.org/10.1017/s1352465820000430>
- Ghaderi, B., Yazdanbakhsh, K., & Karami, J. (2022). Development and testing of a structural model of social anxiety based on maladaptive early schemas and metacognitive beliefs, with the mediating role of cognitive emotion regulation strategies in adolescent girls. *Psychological Sciences Monthly*, 21(109), 123-144. <https://doi.org/10.52547/JPS.21.109.123>
- Hashemi, Z., Aini, S., & Shahjouei, T. (2022). Effectiveness of Metacognitive Therapy Based on Executive Function Training on Cognitive-Attentional Syndrome in Students with Social Anxiety. *Clinical Psychology*, 14(2), 1-11. http://edcbmj.ir/files/site1/user_files_944f39/zohreh-A-10-2348-4-3b83f2e.pdf
- Hashemi, Z., Eyni, S., & Shahjoei, T. (2022). The Effectiveness of Metacognitive Therapy Brain-Based Executive Functioning Training on Cognitive-Attention Syndrome in Students with Social Anxiety. *Education Strategies in Medical Sciences*, 15(2), 122-130. <https://www.sid.ir/paper/1021490/en>
- Nateghi, N., Dadashi, M., & Mahmoud Alilou, M. (2019). Metacognitive Therapy in Improving sign, Metacognitive Beliefs and Thought Fusion in Patients with Obsessive-Compulsive Disorder [Original Research Article]. *Middle Eastern Journal of Disability Studies*---, 9(0), 112-112. <http://jdisabilstud.org/article-1-991-en.html>
- Nematollahi, S., Afshari Nia, K., Alsadat Hosseini, S., Kakabraei, K., & Parandin, S. (2023). The Effectiveness of Metacognitive Therapy and Compassion-Based Therapy on Reducing Rumination, Anxiety and Worry in Obsessive-Compulsive Patients. *Journal of Psychological Dynamics in Mood Disorders (PDMD)*, 2(3), 50-61. <https://doi.org/10.22034/pdmd.2023.188277>
- Olivari, C., Mansueto, G., Marino, C., Candellari, G., Cericola, J., Binnie, J., & Caselli, G. (2025). Metacognitive beliefs and desire thinking as potential maintenance factors of compulsive sexual behavior. *Addictive behaviors*, 161, 108214. <https://doi.org/10.1016/j.addbeh.2024.108214>
- Qarali, Z., & saberi, h. (2020). Relationship between Metacognition and Delusional Beliefs with Obsessive-Compulsive Symptoms. *Thoughts and Behavior in Clinical Psychology*, 15(56), 17-27. https://jtbcp.roudehen.iau.ir/article_1854_eb87d9ae095cceeab1c5afed396eaa56.pdf
- Rabat Mili, S., & Karimi, M. (2018). The predictive role of metacognitive beliefs, mindfulness, and fear of negative evaluation in social anxiety among adolescents. *Social Psychology Research*, 8(31), 51-68. https://www.socialpsychology.ir/article_87485.html?lang=en
- Rajabi, M. (2018). *Comparing the effectiveness of metacognitive therapy, Acceptance and Commitment Therapy, and problem-solving therapy on the symptoms of social anxiety, quality of life, and feelings of loneliness* Master's thesis in Psychology, Lorestan University].
- Reshadat, S., Morovati, F., Zakiee, A., & Rajabi Gilan, N. (2019). Predicting social anxiety based on components of

- mindfulness, experiential avoidance, and metacognitive beliefs among adolescents. *Journal of Mazandaran University of Medical Sciences*, 29(176), 92-104. <https://jmums.mazums.ac.ir/article-1-13162-en.html>
- Samia Shahani, S. S., Ahmadi, O., Nemati, A., & Hajizade Sanzighi, A. (2020). The mediating role of metacognitive beliefs in the relationship between the cognitive order of emotion and mindfulness. *Applied Research in Consulting*, 3(1), 83-103. https://www.jarci.ir/article_704282_bfae31d3c8978e5458e0dddf6fb89dc2.pdf
- Strand, E. R., Nordahl, H., Hjemdal, O., & Nordahl, H. M. (2023). Metacognitive Beliefs Predict Interpersonal Problems in Patients With Social Anxiety Disorder. *Scandinavian journal of psychology*, 64(6), 819-824. <https://doi.org/10.1111/sjop.12943>
- Williams, M., Honan, C., & Matthews, A. J. (2024). Attentional bias to threat: an investigation of psychological predictors beyond trait anxiety. *Current Psychology*. <https://doi.org/10.1007/s12144-024-05641-1>
- Xu, Y., Zheng, P., Feng, W., Chen, L., Sun, S., Liu, J., Tang, W., Bao, C., Xu, L., Xu, D., & Zhao, K. (2023). Patterns of Attentional Bias in Antenatal Depression: An Eye-Tracking Study. *Frontiers in Behavioral Neuroscience*, 17. <https://doi.org/10.3389/fnbeh.2023.1288616>
- Zhang, F., Huang, C. W., Yan, W., Ouyang, H., & Liu, W. (2024). Attentional Bias Modification and Attention Control Training in PTSD: A Systematic Review and Meta-Analysis. *Therapeutic Advances in Psychopharmacology*, 14. <https://doi.org/10.1177/20451253241243260>
- Zhu, S., Zhao, Y., Lu, H., Guo, F., Lin, C., Qiu, H., & Jiang, K. (2024). Examining the Role of Negative Attentional Bias in Patients With Adolescent Depressive Disorders: An Emotion-Attention Blink Task Study. <https://doi.org/10.21203/rs.3.rs-3865806/v1>