




## Predicting Somatic Symptom Disorder Using Ensemble Learning on Personality, Stress, and Emotion Regulation Data

Sinokuthokoza. Segun Oyeyiola<sup>1\*</sup> 

<sup>1</sup> Department of Career Guidance and Counselling, University of KwaZulu-Natal, Durban, South Africa

\* Corresponding author email address: segunoyeyiola@gmail.com

E d i t o r	R e v i e w e r s
Bahram Jowkar  Professor of Psychology Department, Shiraz University, Iran jowkar@shirazu.ac.ir	<b>Reviewer 1:</b> Maryam Aslzaker  Assistant professor of clinical psychology, clinical Psychology Department, Shahid Beheshti University of medical sciences, Tehran, Iran.m.aslzaker@sbm.ac.ir <b>Reviewer 2:</b> Seyed Milad Saadati  Department of Laboratory Medicine and Pathobiology (LMP), University of Toronto, Canada.m.saadati@mail.utoronto.ca

### 1. Round 1

#### 1.1. Reviewer 1

Reviewer:

The sentence “their interactive and nonlinear effects remain insufficiently modeled using traditional statistical frameworks” is conceptually important, but currently assertive rather than demonstrative. Consider briefly specifying which nonlinearities or interaction patterns are theoretically expected (e.g., threshold effects of stress, moderation by emotion regulation) to strengthen the rationale for ensemble learning.

In the paragraph starting “A substantial body of evidence indicates that emotional dysregulation constitutes a central psychopathological dimension...”, the construct of emotion regulation is treated broadly. Please clarify whether the study adopts a process-oriented, strategy-based, or deficit-oriented conceptualization, and how this maps onto the specific subscales used later.

The paragraph beginning “Despite these methodological advances...” would benefit from a more explicit statement of novelty, such as whether this is the first SSD study using stacking ensembles, incremental block modeling, or a South African sample.

The phrase “multi-stage convenience sampling strategy” requires further clarification. Please specify how stages were defined, how participants were distributed across recruitment channels, and whether any steps were taken to reduce selection bias.

The sentence “The final sample size was determined based on recommendations for machine learning applications” is vague. Please report the actual sample size, subject-to-feature ratio, and any a priori rationale for adequacy in ensemble modeling.

The description “a validated somatic symptom measure” is insufficiently specific. Please name the instrument, report its scoring range, and clarify whether it assesses DSM-aligned SSD criteria or general somatic distress.

The sentence “minor linguistic adjustments were made to ensure cultural clarity” requires elaboration. Please specify what adjustments were made, whether expert review or pilot testing was conducted, and how measurement equivalence was preserved.

Authors revised the manuscript and uploaded the document.

## 1.2. Reviewer 2

Reviewer:

The statement “traits such as conscientiousness and emotional stability may serve protective roles” would benefit from greater theoretical precision. Please specify the hypothesized mechanisms (e.g., planning, impulse control, stress appraisal) through which these traits are expected to influence somatic symptoms, particularly in a machine-learning context.

The paragraph beginning “Stress exposure represents another critical factor...” discusses stress broadly. Please clarify whether perceived stress is modeled as a proximal psychological appraisal or as a proxy for chronic environmental load, as this affects interpretability of feature importance results.

In the sentence “Emotion regulation processes mediate the relationship between stress exposure and psychological distress...”, the term mediate implies causal ordering. Given the cross-sectional design, please soften this language or explicitly state that mediation is conceptual rather than empirically tested.

While the paragraph on fronto-limbic networks is informative, the manuscript would benefit from a clearer linkage between neurobiological evidence and the chosen predictors. Please explicitly state how these neural findings justify the inclusion of personality, stress, and emotion regulation variables in predictive models.

The sentence “Machine learning approaches... offer substantial advantages over traditional regression techniques” could be strengthened by briefly acknowledging trade-offs, such as interpretability versus accuracy, to present a more balanced methodological rationale.

Authors revised the manuscript and uploaded the document.

## 2. Revised

Editor’s decision: Accepted.

Editor in Chief’s decision: Accepted.