



Machine Learning Algorithms to Explore the Relationship Between Primary Obsessive-Compulsive Disorder (OCD) and Co-morbid Depression Symptoms

Ajay Kumar Singh
Sabyasachi Mukherjee



Agenda

- The weight of repetition — A Mind Caught Between Knowledge and Fear
- Overview of Obsessive-Compulsive Disorder & Depression
- Clinical Significance and Burden
- Data
- Motivation and Methodology
- Result
- Discussion
- Conclusion

Let's start with a Story - The Weight of Repetition

Introduction

Dr. Arjun Sen, a psychiatrist, spent his career studying the human mind, yet his own thoughts often betrayed him

Every morning before leaving home, he locked his apartment door, then checked it repeatedly

He knew it was secure. Still, a voice whispered, What if you missed something?

Increased distress

At first, he dismissed it as stress

But as his work demanded precision – no error in diagnosing mental illness, the rituals began to consume hours

Each gave him fleeting relief, quickly replaced by a heavier wave of doubt.

His mind behaved both as the prison and the warden.



The Story continues....

Inherited Anxiety

His mother's similar patterns-washing her hands till her skin cracked, terrified of invisible contamination

Mental health struggles can echo across generations, intertwining biology and experience.

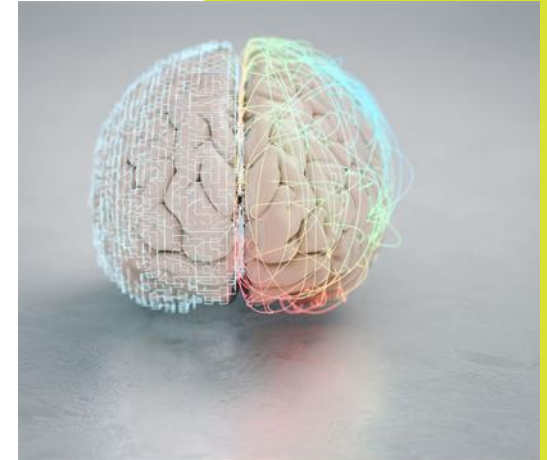
Environmental Triggers & Cognitive patterns

Stressful environments, sleep deprivation, and caffeine use act as triggers that intensify OCD symptoms and compulsive behaviors.

Maladaptive thoughts like intolerance of uncertainty and exaggerated responsibility perpetuate compulsive rituals

Path to Recovery

Through therapy and support, Arjun learned to live with uncertainty, finding peace not in silence but in acceptance.



Overview of Obsessive-Compulsive Disorder & Depression

Understanding prevalence, comorbidity, and clinical relevance



Defining OCD and Depression

Obsessive–Compulsive Disorder (OCD) is characterized by intrusive thoughts and repetitive behaviors, while depression involves persistent low mood, anhedonia, and cognitive impairment. Both are chronic and disabling psychiatric conditions.



Clinical Overlap and Diagnostic Complexity

OCD and depression often share overlapping symptom dimensions—such as guilt, rumination, and indecision—complicating differential diagnosis and treatment planning.



Prevalence and Global Burden

Globally, OCD affects ~2–3% of the population, and depression affects more than 300 million individuals. Their comorbidity rates exceed 40%, amplifying functional impairment and healthcare costs.



Impact on Function and Quality of Life

Comorbid cases experience greater symptom severity, poorer response to treatment, and higher relapse rates, underscoring the need for integrated diagnostic frameworks.

Clinical Significance & Burden

Functional, economic, and societal impacts



Functional Impairment

Comorbid OCD and depression markedly reduce occupational performance, social engagement, and daily functioning. Individuals often experience severe fatigue, diminished concentration, and social withdrawal.



Public Health Relevance

Depression is a leading cause of disability worldwide, and OCD contributes significantly to the global neuropsychiatric burden. Their intersection magnifies suicide risk and chronicity..



Economic and Healthcare Costs

Global costs exceed \$1 trillion annually due to lost productivity, disability, and treatment demands. Patients with comorbid disorders utilize 40% more mental health services than those with single diagnoses



Need for Integrated Care

Fragmented mental healthcare systems often treat OCD and depression separately, despite overlapping neurobiological and psychosocial mechanisms requiring cross-disciplinary intervention.



Data

Data

- ❖ This analysis is based on secondary data with symptoms of OCD and depression in 408 adult patients with a primary diagnosis of OCD.
- ❖ This data Collected at OCD Center and Cognitive–Behavioral Therapy Services, Rogers Memorial Hospital, USA.
- ❖ The scales used in this data collection were the Quick Inventory of Depressive Symptomatology - self-report version (QIDS-SR), and the Yale-Brown Obsessive Compulsive Scale - self-report (Y-BOCS-SR).
- ❖ The depression scale has 16 parameters , the OCD scale has 10 parameters



Analysis Data

Description of features measuring obsession and compulsion respectively

VARIABLE NAME	ANALYSIS VARIABLE	DESCRIPTION
Obtime	obtime	Time occupied by obsessive thoughts
Obinterfer	obinterfer	Interference in one's life due to obsessive thoughts
Obdistress	obdistress	Distress associated with obsessive thoughts
Obresist	obresist	Difficulty resisting obsessions
Obcontrol	obcontrol	Difficulty controlling obsessions
Comptime	comptime	Time spent performing compulsive behaviors
Compinterf	compinterf	Interference in one's life due to compulsive behaviors
Compdis	compdis	Distress associated with compulsive behaviors
Compresis	compresis	Difficulty resisting compulsions
Compcont	compcont	Difficulty controlling compulsions

Description of features measuring depression

VARIABLE NAME	ANALYSIS VARIABLE	DESCRIPTION
Onset	Sl_on	Difficulty falling asleep
Middle	SL_mid	Sleep during the night
Late	SL_la	Waking up too early
Hypersom	Hypersom	Sleeping too much
Sad	Sad	Feeling sad
Decappetite	AP_De	Decreased appetite
Incappetite	AP_In	Increased appetite
Weightloss	We_loss	Decreased weight within the last two weeks
Weightgain	We_gain	Increased weight within the last two weeks
Concen	Con	Impairment in concentration/decision-making
Guilt	Guilt	Self-blaming view of oneself
Suicide	Suicide	Suicidal thoughts, plans or attempts
Anhedonia	Anhedonia	Anhedonia
Fatigue	Fatigue	Fatigue
Retard	Retard	Psychomotor retardation
Agitation	Agitation	Agitation



Methodology

Motivation & Methodology

□ Motivation

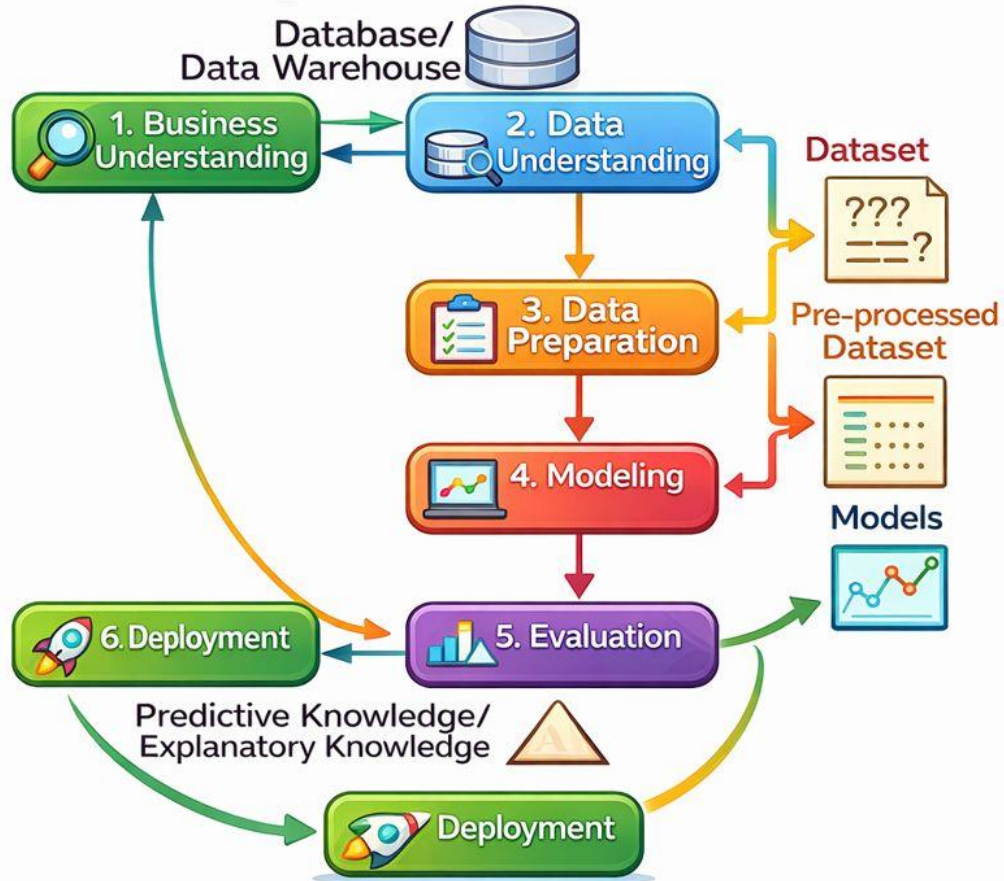
- ❖ The principal motivation of this presentation is twofold: one is to explore different predictive functional relationships between (i) Obsession and depression parameters (ii) Compulsion and depression parameters (iii) together OCD and depression; another one is to develop the understanding and implication of ML and data mining in psychopathology.

□ Methodology

- ❖ Take an average scores for each patient obtained from the questionnaire measuring OCD symptoms to get an average/mean unbiased score of obsession, compulsion and OCD.
- ❖ These average scores for obsession, compulsion and OCD were used as the response/target variable to find the best predictive non-linear relationship (using ML and data mining) between the target variable (obsession, compulsion and OCD) and input variables (depression parameters).
- ❖ For important feature selection, sensitivity analysis has been used. Sensitivity analysis is a technique that is used to determine how the values of independent variables value will determine the value of a particular dependent variable under a given set of assumptions.

Methodology

CRISP-DM process model



CRISP-DM Process Model

The CRISP-DM (Cross-Industry Standard Process for Data Mining) model shows the step-by-step process used in data science and data mining projects.

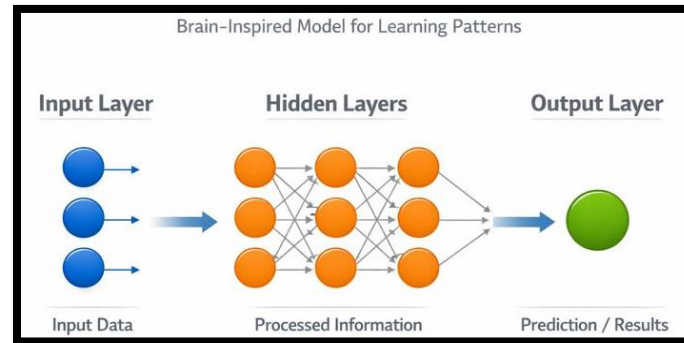
- **Business Understanding** – Define the business problem and goals clearly.
- **Data Understanding** – Collect data and explore it to understand its quality and structure.
- **Data Preparation** – Clean, transform, and organize data for modeling.
- **Modeling** – Apply machine learning or statistical techniques to build models.
- **Evaluation** – Check if the model meets business objectives and performs well.
- **Deployment** – Use the model in real-world applications to generate insights or predictions.

The process is cyclical, meaning you can go back to earlier steps to improve results.

Methods

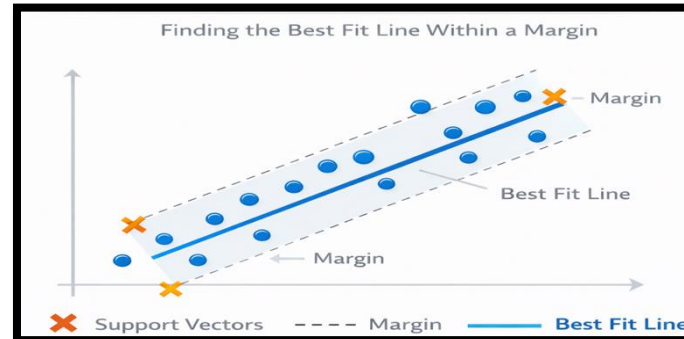
❑ Models (Machine Learning)

❖ Artificial neural network (ANN) model



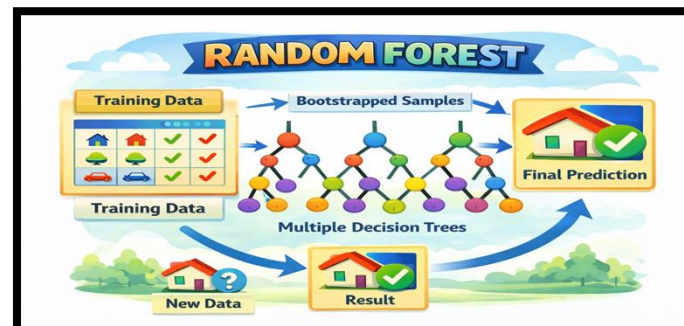
Artificial Neural Network (ANN) is a brain-inspired model that learns patterns by passing data through layers of connected neurons. It takes input data, processes it through hidden layers and produces an output used for prediction or classification.

❖ Support Vector Machine (SVM) model



Support Vector Machine (SVM) predicts continuous values by finding a function that deviates from actual targets by at most a specified margin while keeping the model as flat as possible. It can handle **non-linear relationships** using kernel functions.

❖ Random Forest (RF) model

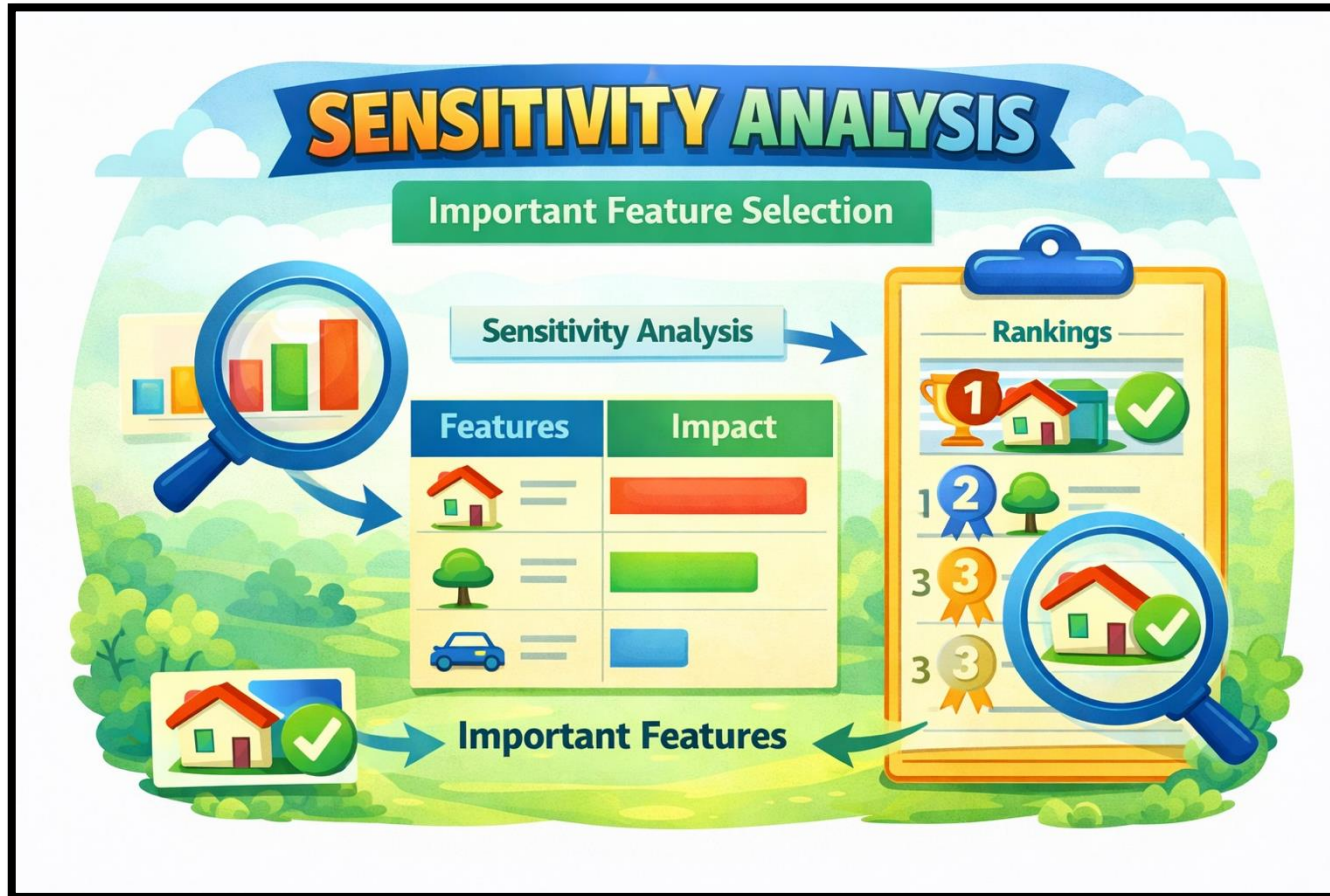


A **Random Forest** is an ensemble learning method for classification or regression. It builds **multiple decision trees** and merges them to get more accurate and stable predictions.

Methods

❑ Feature Selection

❖ Sensitivity Analysis



Sensitivity analysis is a technique used to determine how changes in input features affect a model's output. In feature selection, it helps identify which features are most influential on predictions. Let $y_{a,j}$ denote the output obtained by holding all input variables at their average values except x_a , which varies through its entire range ($x_{a,j}$, with $j \in \{1, 2, \dots, L\}$ levels). The variance (V_a) of $y_{a,j}$ used as a measure of input relevance. A high variance (V_a) suggests a high x_a relevance, thus the input relative importance (R_a) is given by:

$$R_a = \frac{V_a}{\sum_{i=1}^I V_i \times 100(\%)}$$

How it works:

- Each feature is varied slightly while keeping others constant.
- The change in model output is measured.
- Features causing larger output changes are considered more important.

Why it's useful:

- Identifies key features that strongly impact predictions
- Improves model interpretability
- Helps reduce irrelevant or redundant features

Evaluation Metrics for Model Prediction

Performance Evaluation Metrics

Metric	Full Form	Formula	Description
SAE	Sum of Absolute Errors	$\sum y_i - \hat{y}_i $	Total absolute error
MAE	Mean Absolute Error	$\frac{1}{n} \sum y_i - \hat{y}_i $	Average absolute error
MSE	Mean Squared Error	$\frac{1}{n} \sum (y_i - \hat{y}_i)^2$	Average squared error
SSE	Sum of Squared Errors	$\sum (y_i - \hat{y}_i)^2$	Total squared error
RMSE	Root Mean Squared Error	$\sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$	Root of MSE



Results

Obsession Results – Model Comparison (NN vs SVM)

Neural Network (NN)

Run	SAE	MAE	MSE	SSE	RMSE
1	44.58697	0.32784	0.20141	27.39303	0.44879
2	47.86138	0.35192	0.21202	28.83574	0.46046
3	45.43233	0.33406	0.20509	27.89334	0.45287
4	41.10860	0.30226	0.18808	25.57892	0.43368
5	47.52364	0.34943	0.21078	28.66735	0.45911
Mean	45.30258	0.33310	0.20348	27.67368	0.45098
95% CI	± 3.37763	± 0.02483	± 0.01195	± 1.62520	± 0.01336

Support Vector Machine (SVM)

Run	SAE	MAE	MSE	SSE	RMSE
1	43.78497	0.32194	0.18769	25.52613	0.43323
2	45.14387	0.33194	0.17889	24.32959	0.42295
3	47.25219	0.34744	0.19642	26.71385	0.44319
4	46.79813	0.34410	0.18656	25.37252	0.43192
5	45.74196	0.33633	0.19055	25.91608	0.43653
Mean	45.74423	0.33635	0.18802	25.57163	0.43357
95% CI	± 1.70969	± 0.01257	± 0.00791	± 1.07668	± 0.00914

– SVM's better overall accuracy (lower MSE, SSE, RMSE)

– Stability advantages of SVM (narrower confidence intervals)

Compulsion Results – Model Comparison (NN vs SVM)

Neural Network (NN)

Run	SAE	MAE	MSE	SSE	RMSE
1	51.46456	0.37841	0.23414	31.84389	0.48388
2	53.84281	0.39590	0.26657	36.25469	0.51631
3	52.69002	0.38742	0.28896	39.29864	0.53755
4	47.93030	0.35242	0.25808	35.09957	0.50802
5	53.66637	0.39460	0.30172	41.03443	0.54929
Mean	51.91881	0.38175	0.26989	36.70624	0.51901
95% CI	± 3.00736	± 0.02211	± 0.03288	± 4.47178	± 0.03179

- SVM achieves lower Mean SAE, MAE, MSE, SSE, and RMSE compared to NN, indicating superior predictive performance.

Support Vector Machine (SVM)

Run	SAE	MAE	MSE	SSE	RMSE
1	49.49664	0.36394	0.22367	30.41930	0.47293
2	44.61625	0.32806	0.19946	27.12666	0.44661
3	52.52254	0.38619	0.29096	39.57140	0.53941
4	48.63455	0.35760	0.27606	37.54477	0.52541
5	47.10926	0.34639	0.24459	33.26480	0.49456
Mean	48.47585	0.35644	0.24695	33.58539	0.49578
95% CI	± 3.63104	± 0.02669	± 0.04640	± 6.31149	± 0.04696

- SVM demonstrates better average accuracy across most metrics.
- Overall, SVM outperforms NN for the Compulsion dataset.

OCD Results – Model Comparison (NN vs SVM)

Neural Network (NN)

Run	SAE	MAE	MSE	SSE	RMSE
1	56.64063	0.41648	0.29376	39.95120	0.54200
2	54.41534	0.40011	0.25276	34.37592	0.50276
3	51.15987	0.37618	0.21654	29.45002	0.46534
4	63.41354	0.46628	0.34535	46.96785	0.58767
5	62.69511	0.46099	0.33258	45.23145	0.57670
Mean	57.66490	0.42401	0.28820	39.19529	0.53489
95% CI	± 6.57830	± 0.04837	± 0.06703	± 9.11621	± 0.06347

- SVM achieves lower Mean SAE, MAE, MSE, SSE, and RMSE compared to NN
- NN exhibits higher variability, reflected by wider confidence intervals

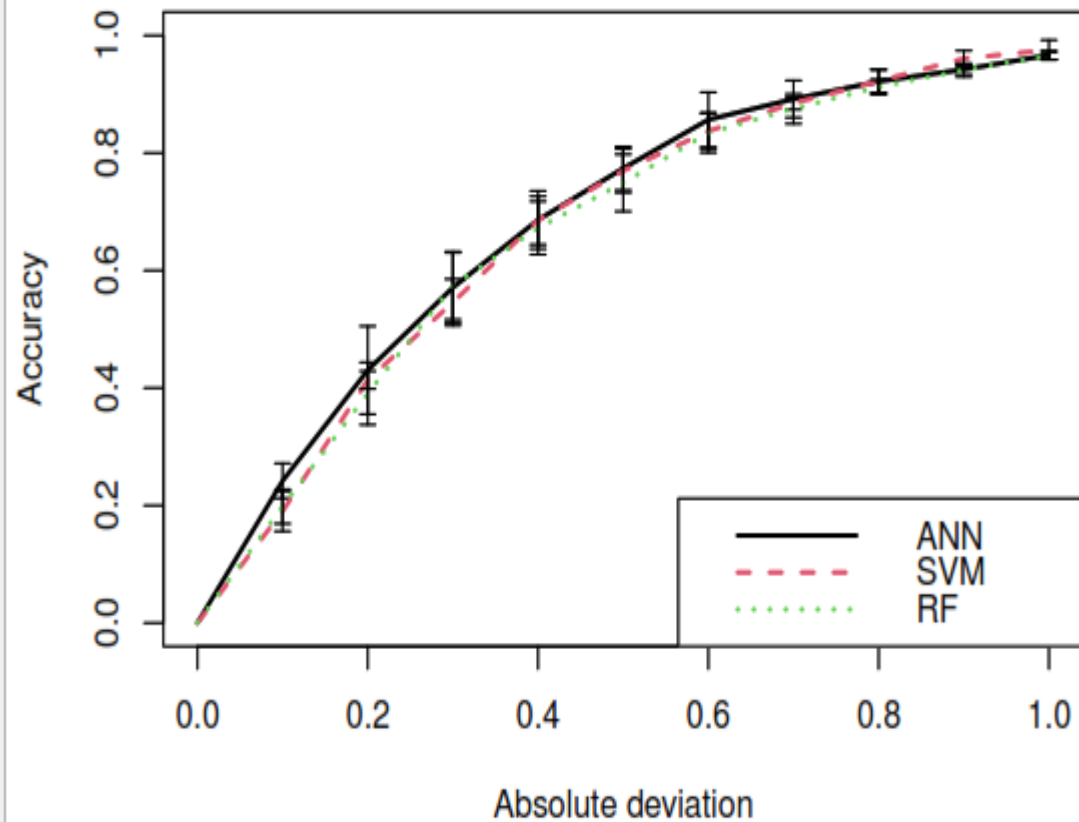
Support Vector Machine (SVM)

Run	SAE	MAE	MSE	SSE	RMSE
1	54.22839	0.39874	0.27900	37.94331	0.52820
2	53.87476	0.39614	0.23949	32.57030	0.48937
3	53.60938	0.39419	0.25247	34.33593	0.50246
4	56.96495	0.41886	0.27753	37.74375	0.52681
5	53.65952	0.39456	0.25063	34.08609	0.50063
Mean	54.46740	0.40050	0.25982	35.33588	0.50950
95% CI	± 1.75981	± 0.01294	± 0.02180	± 2.96488	± 0.02135

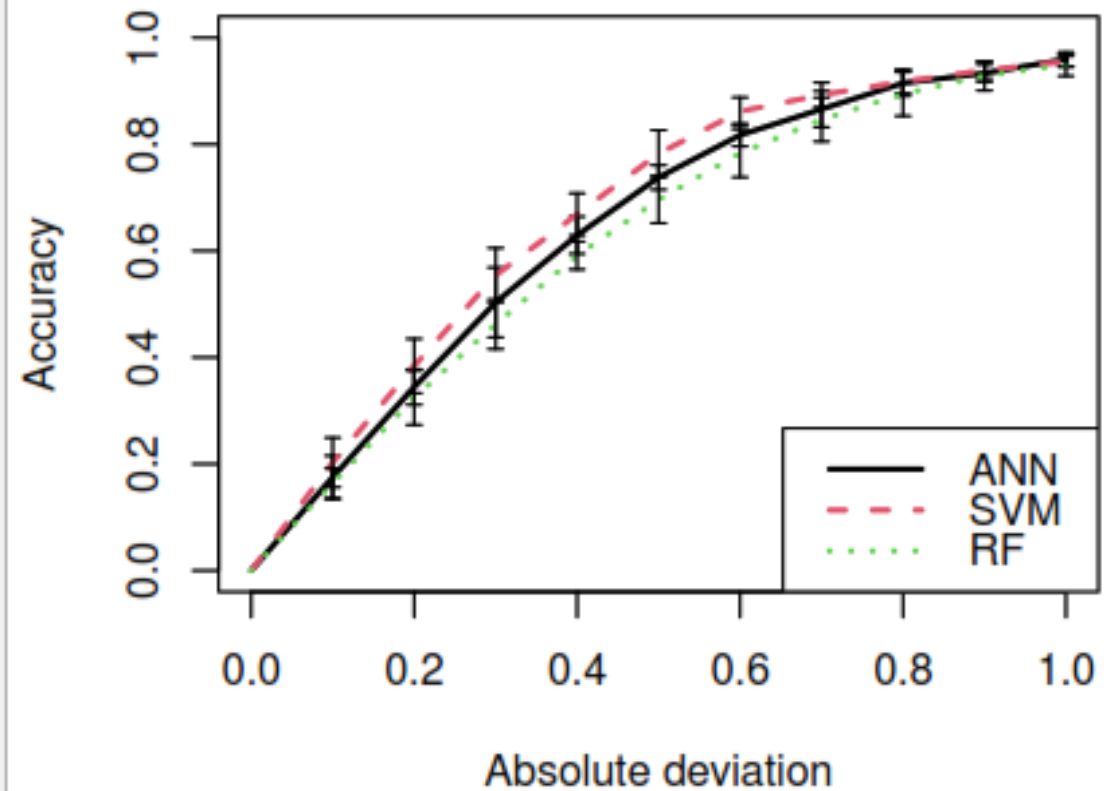
- SVM shows more stable and accurate performance for OCD prediction.
- Overall, SVM outperforms NN across all evaluated error metrics

Overall Accuracy Comparison (all three models)

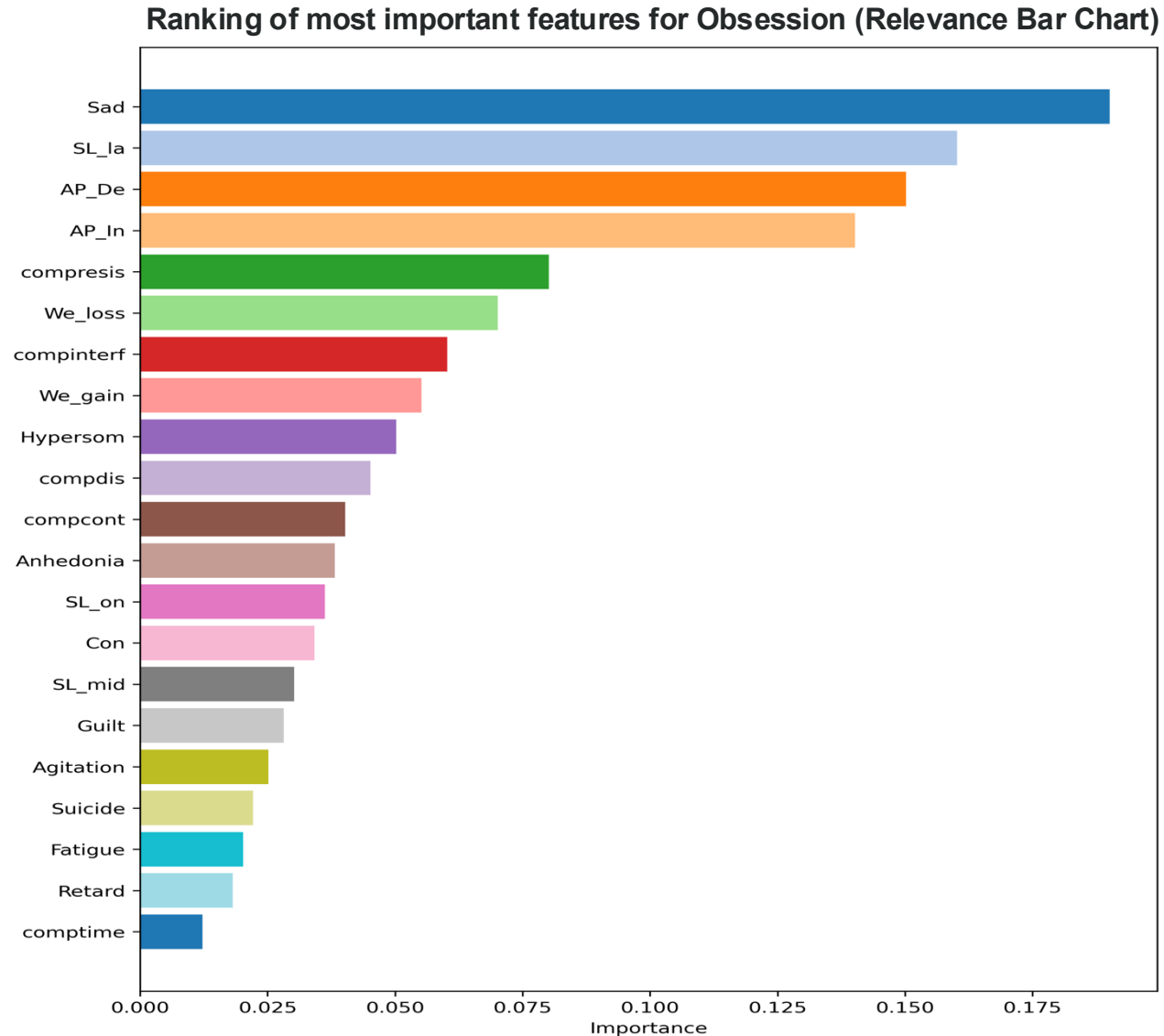
Accuary Comparison for Obsession level



Accuary Comparison for Compulsion level



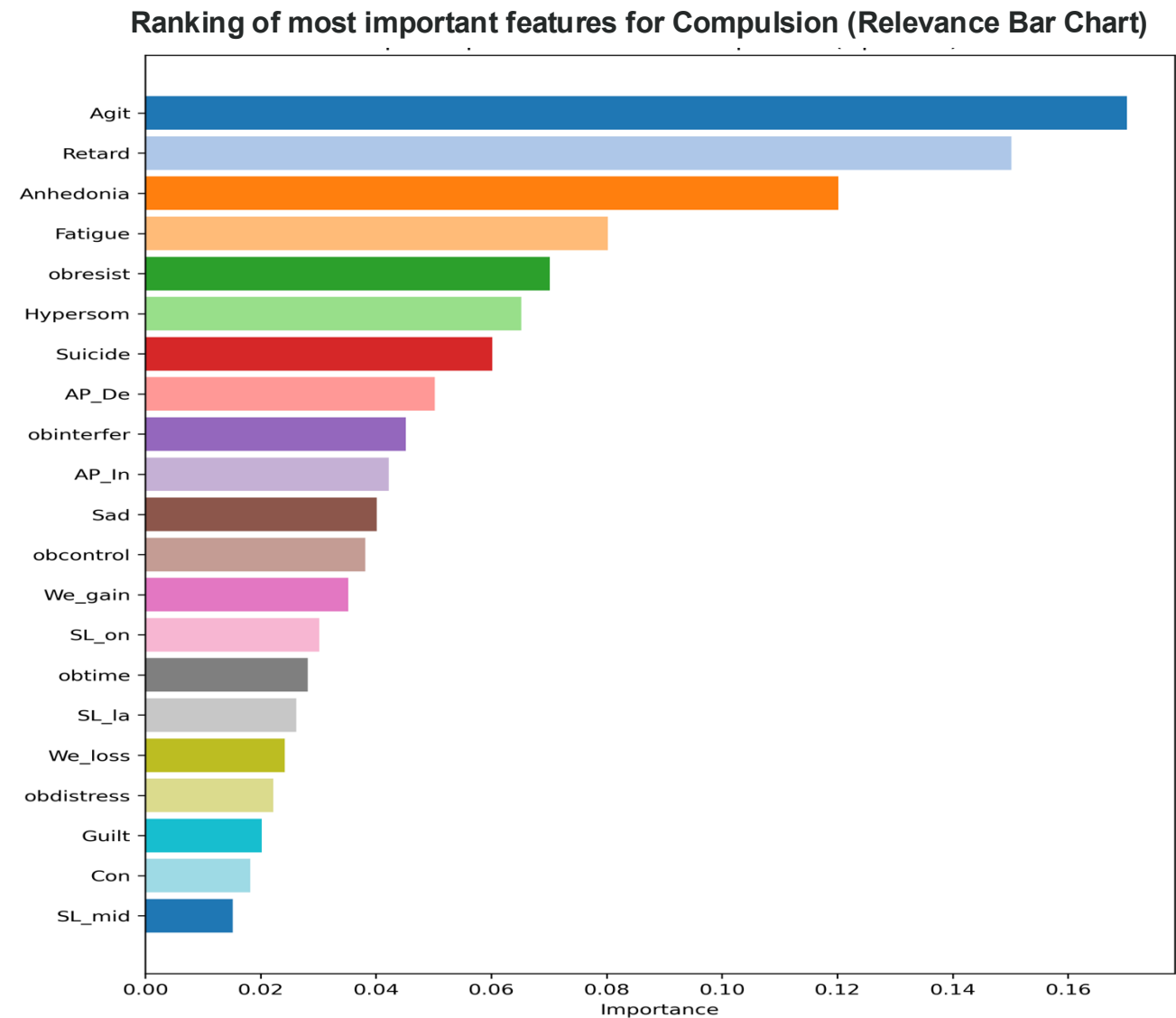
Feature Selection using Sensitivity Analysis (under best model SVM) - Obsession



Discussion on Relevance Bar Chart for Obsession

- Obsessive symptoms showed strongest associations with affective and physiological factors, especially depressed mood, sleep latency, and appetite disturbances.
- Sadness emerged as the most influential predictor, underscoring the close link between obsessions and depressive affect.
- Findings align with Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), American Psychiatric Association (APA) models of Obsessive–Compulsive Disorder (OCD) and Major Depressive Disorder (MDD), which frequently co-occur.
- The strong obsession–sadness link supports evidence that comorbid depression increases obsessional severity and functional impairment.
- Sleep latency was highly associated with obsession, suggesting intrusive thoughts and cognitive hyperarousal disrupt sleep initiation.
- Sleep disturbance, common in both OCD and depression, may further impair emotional regulation and reinforce obsessive thinking.
- Both increased and decreased appetite were linked to obsession, reflecting heterogeneous stress and coping responses.
- Appetite disturbance may indicate neurovegetative dysregulation or maladaptive coping (e.g., emotional eating).
- Overall, mood and neurovegetative symptoms outweighed compulsive behaviors, supporting a transdiagnostic framework.
- Clinically, targeting mood and sleep disturbances may indirectly reduce obsessive symptom severity.

Feature Selection using Sensitivity Analysis (under best model SVM) - Compulsion

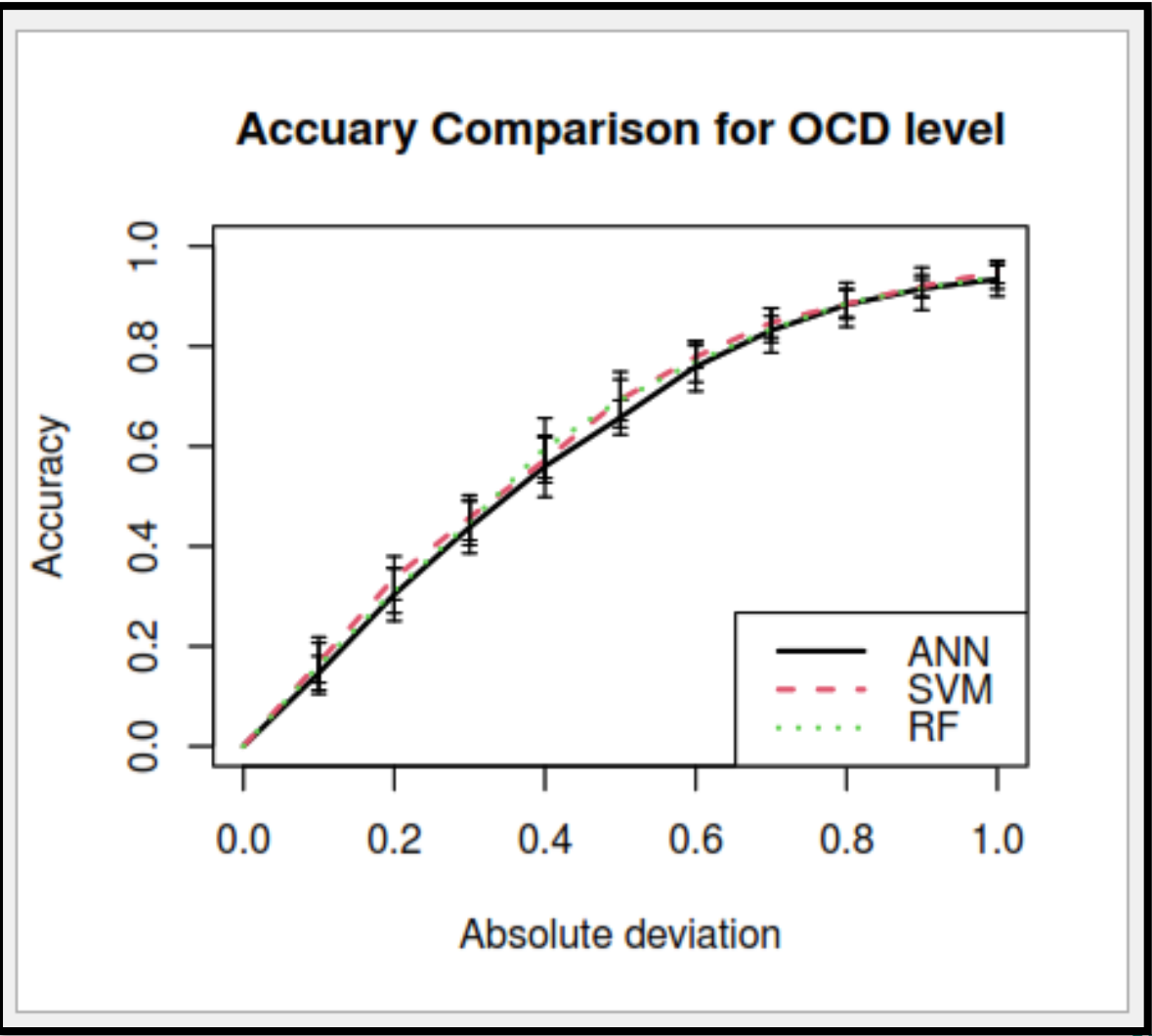
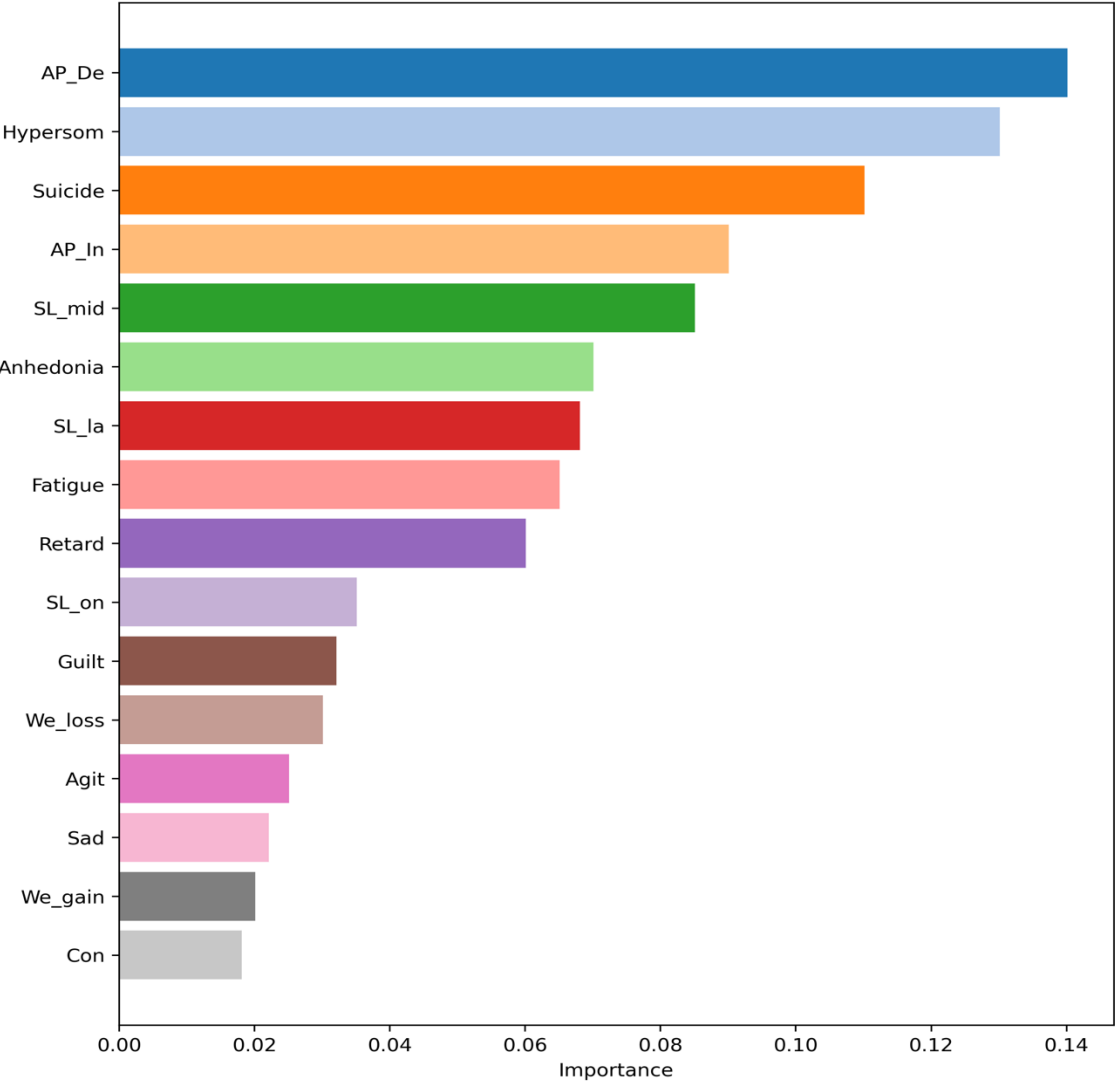


Discussion on Relevance Bar Chart for Compulsion

- Compulsive symptoms were most strongly associated with psychomotor agitation, psychomotor retardation, anhedonia, and fatigue.
- This pattern suggests compulsions are closely linked to motor activity and energy dysregulation, rather than affective distress alone.
- Psychomotor agitation emerged as the strongest predictor, consistent with DSM-5 descriptions of compulsions as driven, repetitive behaviors performed to relieve internal tension or anxiety.
- The association with psychomotor retardation indicates heterogeneity in compulsive presentation, with behaviors becoming rigid, effortful, and automatic, overlapping with depressive motor slowing.
- Anhedonia was strongly related to compulsion, supporting the role of negative reinforcement rather than pleasure in maintaining compulsive behaviors.
- This aligns with DSM-5 criteria emphasizing that compulsions are not experienced as enjoyable but are performed to reduce distress or prevent feared outcomes.
- Fatigue reflects the cognitive and physical burden of repetitive rituals and reduced capacity to resist compulsive urges.
- Overall, findings support a DSM-5–based view of compulsions as repetitive behaviors embedded in psychomotor and neurovegetative dysfunction.
- Clinically, assessment and treatment should address motor and energy-related disturbances, especially in the presence of comorbid depressive features.

Feature Selection using Sensitivity Analysis (under best model SVM) – OCD (together)

Ranking of most important features for OCD (Relevance Bar Chart)



Discussion on Relevance Bar Chart for OCD

- SVM achieves lower Mean SAE, MAE, MSE, SSE, and RMSE compared to NN
- NN exhibits higher variability, reflected by wider confidence intervals
- SVM shows more stable and accurate performance for OCD prediction. Overall, SVM outperforms NN across all evaluated error metrics.
- Decreased appetite (AP_De) showed the highest importance, indicating strong influence on OCD symptom severity.
- Hypersomnia emerged as the second most influential factor, highlighting significant sleep-related dysregulation.
- Suicidality ranked third, suggesting severe depressive distress is closely linked to obsessive–compulsive symptoms.
- Increased appetite (AP_In) was also highly influential, reflecting neurovegetative appetite disturbances.
- Middle insomnia (SL_mid) followed closely, reinforcing the role of sleep disruption in OCD.

Overall interpretation:

The most influential predictors are predominantly depressive and neurovegetative symptoms, particularly appetite disturbance, sleep dysregulation, and suicidality, underscoring that obsession–compulsion symptoms are strongly embedded within depressive symptom networks rather than isolated anxiety processes.

Conclusion

This study demonstrates that:

- ML techniques, particularly SVM, effectively model complex relationships between depressive and OCD symptoms.
- Different depressive symptoms predict obsession and compulsion uniquely. For obsession, sadness is relevant when it is persistent
- Affective (sadness), neurovegetative (sleep and appetite), and psychomotor (agitation/retardation) symptoms drive OCD severity.
- Low error & variance for obsession compared to compulsion is showing that symptoms for obsession have better association with depressive symptoms
- Comorbid depression shapes OCD presentation in systematic, predictable ways.

Clinically, the results support the adoption of transdiagnostic treatment approaches addressing mood, sleep, and psychomotor functioning alongside traditional OCD interventions such as ERP (Exposure and Response Prevention).

References

- ❖ American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.; DSM-5-TR). APA Publishing.
- ❖ Abramowitz, J. S., Taylor, S., & McKay, D. (2009). Obsessive–compulsive disorder. *The Lancet*, 374(9688), 491–499. [https://doi.org/10.1016/S0140-6736\(09\)60240-3](https://doi.org/10.1016/S0140-6736(09)60240-3)
- ❖ Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive–compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry*, 15(1), 53–63. <https://doi.org/10.1038/mp.2008.94>
- ❖ <https://www.ncbi.nlm.nih.gov/books/NBK553162/>
- ❖ <https://pmc.ncbi.nlm.nih.gov/articles/PMC5777899/>
- ❖ <https://news.un.org/en/story/2017/03/554462>
- ❖ [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(20\)30432-0/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30432-0/fulltext)

Thanks

- Thank you
- Refer to the paper ML05 for details
- Feel free to contact at Ajaykumar.singh@iconplc.com and Sabyasachi.Mukherjee@iconplc.com for any questions and comments



ICONplc.com

© 2025 ICON. All rights reserved.