RSPC: A Relationship-Aware Benchmark for Modeling Stress and Psychiatric Signals in Digitally Mediated Relationships

***Abstract*—Digital communication increasingly mediates intimate relationships, yet computational mental-health research often models distress at the individual level, detached from relational and technological context. Long-distance romantic relationships (LDRs) provide a setting where psychological stress is closely tied to interactional factors such as communication gaps, commitment uncertainty, and time-zone misalignment. We introduce the *Relational Stress and Psychiatry Corpus* (RSPC), a clinically grounded dataset of 1,799 Reddit posts annotated with (1) multi-label psychiatric symptom categories aligned with DSM-5-TR and ICD-11, (2) relational stressor types, and (3) temporal phases of the LDR cycle. We benchmark transformer-based models on three tasks psychiatric symptom classification, relational trigger detection, and temporal-phase prediction. Results show strong performance on frequent labels but persistent difficulty with clinically specific and minority categories. RSPC establishes the first benchmark for modeling mental health as a relational and interaction-dependent phenomenon, enabling more socially grounded mental-health NLP research.**

***Index Terms*—Mental health, affective computing, social computing, social media analysis, long-distance relationships, clinical natural language processing, DSM-5, ICD-11, temporal modeling, digital psychiatry.**

I. Introduction

Digital communication platforms have become primary mediators of intimate human relationships. Messaging applications, video conferencing tools, and social networking platforms increasingly shaped how individuals experience emotional support, conflict, attachment, and separation. While this transformation has enabled new forms of connection across geographical boundaries, it has also introduced novel patterns of psychological stress that are fundamentally relational and technologically mediated. Despite this shift, most computational mental health research continues to model psychological distress as a static, individual-level phenomenon, largely detached from the relational and digital environments in which symptoms often emerge.

Long-distance romantic relationships (LDRs), in which partners maintain an ongoing romantic relationship while living in geographically separated locations, represent a paradigmatic and rapidly growing instance of digitally mediated relational stress, characterized by chronic physical separation, circadian misalignment, limited non-verbal communication, and heightened uncertainty regarding availability and fidelity. Clinical and psychological studies associate these conditions with elevated rates of insomnia, anxiety disorders, adjustment disorders, attachment dysregulation, and depressive symptomatology. Crucially, these conditions rarely arise in isolation; rather, they emerge through recurring interactional triggers such as prolonged communication silence, delayed responses, jealousy episodes, misaligned time zones, and repeated separation–reunion cycles. These triggers interact dynamically with individual vulnerability, giving rise to temporally evolving symptom trajectories that cannot be captured by static diagnostic labels alone.

However, existing natural language processing (NLP) benchmarks for mental health overwhelmingly reduce this complexity to coarse binary or single-label classifications of depression or suicidality. Prominent datasets such as CLPsych [1], eRisk, [2] and Reddit-derived depression corpora treat distress as an individual and context-agnostic construct, with little to no representation of relational causality, clinically differentiated disorders, or temporal recurrence [3], [4]. This methodological abstraction limits the development of applied systems for relationship-aware counseling, early intervention, and longitudinal risk monitoring that can reason about how relational stressors are associated with clinically-grounded psychiatric categories..

We address this critical methodological gap by introducing the first *clinically grounded, relationship-aware, and temporally structured mental health dataset* specifically designed to model how digitally mediated relational stressors co-occur with clinically-grounded psychiatric categories defined in the DSM-5-TR and ICD-11. This framework generalizes beyond LDRs to broader web-mediated relational domains, such as remote caregiving and online dating, and supports future research on relationship-aware language models for assistive, human-in-the-loop counselling analysis.

Our contributions are:

- We introduce RSPC, the first dataset that links psychiatric symptom categories, relational stressors, and temporal phase in digitally mediated romantic relationships.
- We design a clinically-motivated annotation schema aligned with DSM-5-TR and ICD-11 terminology.
- We benchmark multiple transformer-based models across three supervised tasks and analyze failure modes including comorbidity and label imbalance.
- We release the dataset under a privacy-preserving research license to support responsible computational mental-health research.

By unifying psychiatry, affective computing, and social natural language processing, this work lays the foundation for a new class of context-sensitive, clinically informed, and socially responsible digital mental health systems.

We frame our study around the following research questions:

- **RQ1:** Can transformer-based models infer clinically grounded psychiatric symptom categories from LDR narratives?
- **RQ2:** Do relational stressors provide diagnostic value beyond symptom-only text?
- **RQ3:** Can temporal phase information improve modeling of relational distress?
- **RQ4:** What are the limitations of current NLP models in modeling comorbidity and clinically specific and lower-frequency disorders?

II. Related Work

This work sits at the intersection of computational psychiatry, social computing, and clinical psychology. We review five key areas: social media mental health corpora, temporal modeling, digital relationship analysis, clinical grounding in NLP, and ethical frameworks.

A. Social Media Mental Health Corpora

The field of computational mental health has matured through the release of shared tasks and large-scale datasets. Early benchmarking efforts were driven by the **CLPsych** workshops [1], [5], [6] and **CLEF eRisk** labs [7]–[9], which focused on detecting depression, PTSD, and self-harm from Twitter and Reddit streams. To address data scarcity, researchers developed distantly supervised methods to curate massive datasets like **RSDD** [10] and **SMHD** [11], containing millions of posts from users with self-reported diagnoses.

Recent work has expanded into more specific domains, such as stress detection in **Dreaddit** [12], suicide risk assessment [13], [14], and multi-task learning for comorbidity [15]. However, these resources predominantly model mental health as a static, user-level attribute [16]. They lack the relational context such as interpersonal conflict or separation events that is often the immediate precipitant of psychiatric symptoms in clinical practice.

B. Temporal and Longitudinal Modeling

Recognizing the dynamic nature of mental illness, recent research has moved beyond static classification where , psychological states are inferred from isolated posts or aggregated user histories without temporal context. The **RSDD-Time** dataset [17] and subsequent longitudinal studies [18], [19] utilize temporal timestamps to track symptom progression. The eRisk ”early detection” tasks [20], [21] challenge models to identify risk with minimal text history. Deep learning architectures involving LSTMs, temporal attention, and hierarchical transformers have been widely adopted to capture these dependencies [22], [23].

Despite these advances, existing temporal models track *chronological time* (calendar dates) rather than *event time* (time relative to a stressor). Our work introduces ”event-contingent” modeling, linking symptom trajectories directly to the cyclical separation–reunion phases characteristic of LDRs.

C. Digital Relationships and LDR Dynamics

In parallel, social computing research has examined how digital tools mediate romantic intimacy. Foundational studies have analyzed the linguistic signatures of relationship formation, maintenance, and dissolution on platforms like Facebook and Twitter [24]–[26]. Specific to Long-Distance Relationships (LDRs), psychological research highlights the ”idealization” effect of mediated communication [27], [28] and the stress of ”connected presence” [29], [30].

While these studies provide theoretical grounding, they rely heavily on surveys or network metadata. When text analysis is used, it focuses on sentiment or satisfaction [31], [32], not clinical pathology. There is a lack of annotated corpora that map these relational dynamics to the rigorous psychiatric criteria found in the DSM-5-TR [33] or ICD-11 [34].

D. Clinical Grounding and Domain Adaptation

A major challenge in mental health NLP is the gap between social media language and clinical nosology. General-purpose models like BERT often struggle with the nuances of psychiatric text.

However, ”clinical grounding” involves more than just pretraining; it requires valid annotation schemas. Recent critiques have noted that many NLP studies use ad-hoc labels that do not align with diagnostic manuals [4], [35]. Our dataset addresses this by strictly adhering to DSM-5-TR criteria describing conditions such as Separation Anxiety Disorder (SAD) and Adjustment Disorder, which are under-represented in NLP literature despite their prevalence in relational contexts [36], [37].

E. Ethical Frameworks in Digital Psychiatry

Finally, the sensitive nature of this data necessitates strict ethical adherence. We draw upon established frameworks for social media health research [38], [39], which emphasize privacy preservation beyond simple anonymization. The risk of ”predictive policing” or unconsented surveillance in mental health AI is a growing concern [40], [41]. Our work adopts a non-interventionist, research-only release policy, aligning with recent guidelines on responsible AI in healthcare [42], [43].

III. Methodology: The Relational Stress and Psychiatry Corpus (RSPC)

We introduce the **Relational Stress and Psychiatry Corpus (RSPC)**, a novel benchmark designed to bridge the gap between social NLP and clinical psychiatry. Unlike prior datasets that aggregate generic distress signals, RSPC is constructed to capture the relational and contextual factors associated with expressions of psychological distress within the specific context of Long-Distance Romantic Relationships (LDRs). The construction process followed a three-phase pipeline: (1) Domain-Specific Data Curation, (2) Psychiatrist-Guided Annotation Schema Design, and (3) Multi-Stage Annotation and Adjudication.

A. Phase 1: Domain-Specific Data Curation

We targeted the unique communicative environment of LDRs, where relationship maintenance is entirely dependent on digital mediation.

1) Source Selection: Posts were manually collected from publicly available Reddit communitiesfocusing on high-density communities including `r/LongDistance`, `r/LDR`. These forums contain first-person narratives in which users describe relationship stressors and associated psychological experiences in their own words.

2) Filtering and Preprocessing: To improve signal quality, posts were retained only if they contained:

- **Narrative Depth:** sufficient narrative content for contextual interpretation (short posts were excluded).
- **Relational-Clinical Intersection:** description of both relational circumstances and associated emotional or psychological responses.
- **Active Interaction:** Only posts with at least 2 comments were retained to capture community validation of the distress.

Labels reflect symptom descriptions inferred from user-generated language rather than formal diagnosis. All text was anonymized by replacing usernames and proper names with tokens (e.g., `[USER]`), ensuring privacy compliance.

B. Phase 2: The Clinically Grounded Annotation Schema

Our schema departs from lay definitions of "sadness" or "stress." Instead, it is grounded in the diagnostic criteria of the **DSM-5-TR** [33] and **ICD-11** [34]. We developed a hierarchical coding manual in collaboration with a licensed psychiatrist.

1) Tier A: Psychiatric Disorder Labels (Multi-Label): Annotators assessed whether the narrative provided evidence meeting the core symptom criteria for the following disorders:

1) **Major Depressive Disorder (MDD):** Evidence of persistent low mood, anhedonia, worthlessness, or psychomotor agitation lasting > 2 weeks.
2) **Generalized Anxiety Disorder (GAD):** Excessive, difficult-to-control worry about future events, accompanied by physical symptoms (e.g., restlessness, fatigue).
3) **Separation Anxiety Disorder (SAD):** Developmentally inappropriate and excessive fear concerning separation from attachment figures, manifesting as nightmares, distress during travel, or somatic symptoms. Crucial for LDRs.
4) **Adjustment Disorder:** Emotional or behavioral symptoms in response to an identifiable stressor (e.g., a breakup or move) occurring within 3 months of the stressor.
5) **Insomnia Disorder:** Dissatisfaction with sleep quantity/quality (difficulty initiating or maintaining sleep) causing clinically significant distress.

2) Tier B: Relational Trigger Labels: In addition to psychiatric categories, each post was annotated for relational or communication stressors described by the author. These labels capture the interactional contexts that users associate with their distress; they do not imply verified clinical causality. The trigger categories are:

- **Silence Gaps / Ghosting:** Reports of unexpectedly reduced or absent communication, including being "left on read."
- **Jealousy / Insecurity:** Expressions of worry, mistrust, or perceived threat from third-party interactions.
- **Time-Zone Misalignment:** Strain associated with geographically driven differences in waking and communication schedules.
- **Lack of Communication:** Persistent or recurring communication difficulty not necessarily limited to silence events.
- **Reunion / Separation Stress:** Emotional strain surrounding periods of physical reunion or separation.
- **Trust / Fidelity Issues:** Concerns about exclusivity, honesty, or partner reliability.
- **Ambiguity about Commitment:** Uncertainty about the status, future, or seriousness of the relationship.
- **Social Media Surveillance:** Monitoring or anxiety related to partner activity on digital platforms (e.g., online status or posting behavior).

3) Tier C: Temporal Phase (Longitudinal Context): We labeled the position of the user in the LDR cycle: **Separation (Anticipatory)**, **Separation Peak (Acute)**, **Post-Reunion**, or **Breakup/Termination**.

C. Dataset Composition and Statistical Properties

The final RSPC corpus comprises **1,799 annotated posts** collected from Reddit LDR communities and Table I summarizes the key distributional properties.

TABLE I
RSPC DATASET STATISTICS

Property	Count/Percentage
Total Samples	1,799
Multi-label Samples	1,331 (73.99%)
Train/Val/Test Split	1259 / 179 / 361
Disorder Distribution	
Adjustment Disorder (ADJ)	1,341 (74.5%)
Generalized Anxiety (GAD)	1,279 (71.1%)
Separation Anxiety (SAD)	783 (43.5%)
Major Depressive Disorder (MDD)	306 (17.0%)
Insomnia	21 (1.2%)
Temporal Phase Distribution	
Separation (Acute)	1,172 (65.1%)
Anticipation (Pre-separation)	322 (17.9%)
Reunion (Post-reunion)	234 (13.0%)
Unknown	71 (3.9%)

1) Label Imbalance and Clinical Realism: The distribution reflects real-world expected prevalence patterns reported in psychological literature. Adjustment Disorder and GAD are the most frequent diagnoses, consistent with psychological literature showing that relationship stressors typically manifest as situational distress and generalized worry [36]. The severe underrepresentation of Insomnia (1.2%) poses a classification challenge but mirrors its role as a secondary symptom rather than a standalone diagnosis in this context.

2) Co-morbidity Patterns: Multi-label annotations account for 73.99% of the dataset, with the most common co-occurring pairs being ADJ+GAD (n=1,040) and ADJ+SAD (n=696). This high co-morbidity rate supports the design choice of a multi-label formulation and reflects the clinical reality that relationship distress rarely presents as a single isolated disorder.

3) Relational Trigger Landscape: Table II shows that **Commitment Ambiguity** (n=1,191) and **Lack of Communication** (n=1,110) dominate the trigger space, while digital-specific stressors like **Timezone Mismatch** (n=29) and **Social Media Surveillance** (n=37) are comparatively rare despite their clinical salience in LDR literature [30].

TABLE II
RELATIONAL TRIGGER DISTRIBUTION

Trigger Type	Frequency
Commitment Ambiguity	1,191
Lack of Communication	1,110
Reunion/Separation Stress	389
Trust/Fidelity Issues	329
Jealousy/Insecurity	324
Silence Gap	102
Social Media Surveillance	37
Timezone Mismatch	29

IV. TASK DEFINITION

Our dataset supports three supervised NLP tasks: (1) multi-label disorder classification, (2) relational trigger detection, and (3) temporal phase classification. Each task operates over the same input: a Reddit post x, written by an individual describing experiences within a long-distance romantic relationship.

Each post is annotated with: (i) one or more psychiatric symptom categories, (ii) one or more relationship stressor labels, and (iii) a single temporal phase label (where applicable). This structure enables both clinical signal detection and contextual interpretation.

A. Task 1: Multi-Label Disorder Classification

The first task predicts the set of psychiatric disorder categories referenced or implied in the post. Because a single narrative may reflect multiple co-occurring conditions (e.g., anxiety and depression), this is framed as a multi-label classification problem.

Models output a probability score for each disorder category, and a post may receive multiple positive labels. Performance is evaluated using Micro-F1, Macro-F1.

B. Task 2: Relational Trigger Detection

The second task identifies the relationship stressors described by the user, such as ambiguity about commitment, lack of communication, or jealousy/insecurity.

This is also a multi-label task: a post may contain several simultaneous triggers. The aim is to evaluate whether models can infer why distress occurs, rather than only identifying the symptoms themselves.

Evaluation again uses Micro-F1 and Macro-F1.

C. Task 3: Temporal Phase Classification

The third task predicts the temporal phase of the relationship referenced in the post (e.g., anticipation of separation, active separation, post-reunion, or termination).

Unlike the previous tasks, this is a single-label classification problem, where each post is assigned one most-relevant temporal category.

Performance is measured using accuracy and F1-based metrics.

D. Multi-Task Learning Potential

Although we benchmark each task independently, the dataset is designed to support multi-task learning and joint modelling settings where symptom categories, relationship triggers, and temporal context are predicted together. This opens opportunities for modelling interactions between clinical language and relational dynamics in digitally mediated relationships.

V. EXPERIMENTAL RESULTS

All models were evaluated on the RSPC dataset using the 70:10:20 train–validation–test split described earlier. We report Precision, Recall, Macro–F1 and Micro–F1, which are standard metrics for multi–label classification. Micro–averaged scores reflect instance–level correctness, while macro averaging better highlights minority label performance.

A. Task 1: Multi-Label Disorder Classification (MLDC)

We evaluate multi-label disorder classification using standard multi-label evaluation metrics. Table III reports the overall performance on the held-out test set, including macro- and micro-averaged precision, recall, and F1 scores.

TABLE III
TASK 1 RESULTS (DISORDER CLASSIFICATION).

Metric	Score
Macro Precision	0.388
Macro Recall	0.542
Macro F1	0.452
Micro Precision	0.745
Micro Recall	0.721
Micro F1	0.721

To better understand label-wise behavior, we additionally report per-class precision and recall in Table IV. The results indicate strong recall for high-frequency disorders such as Adjustment Disorder (ADJ) and Generalized Anxiety Disorder (GAD), while performance degrades substantially for low-resource labels such as Insomnia.

Overall, these metrics highlight the impact of label imbalance in the dataset: while common anxiety-related disorders are recovered reliably, rare conditions remain challenging. This motivates future work on imbalance-aware learning and data augmentation strategies.

Table V compares representative baseline models for Task 1. The TF–IDF linear classifier provides a strong non-neural

TABLE IV
TASK 1 PER-LABEL PERFORMANCE.

Label	Precision	Recall
SAD	0.50	0.72
ADJ	0.73	1.00
GAD	0.71	0.99
MDD	0.00	0.00
Insomnia	0.00	0.00

TABLE V
TASK 1 (MULTI-LABEL DISORDER CLASSIFICATION) RESULTS ON THE RSPC TEST SET. WE REPORT MACRO-F1 AND MICRO-F1. BEST RESULTS ARE SHOWN IN BOLD.

Model	**Macro-F1**	**Micro-F1**
TF-IDF (Linear)	0.3836	0.7211
BERT-base (uncased)	0.4397	0.7209
ClinicalBERT	0.4253	0.7237
RoBERTa-base	**0.4518**	**0.7451**

baseline, highlighting the role of surface lexical cues in disorder identification. Transformer-based models consistently improve macro-F1, indicating better handling of label imbalance and minority classes. Among them, RoBERTa-base achieves the best overall performance, suggesting that stronger pretraining and contextual representations are beneficial for clinically grounded multi-label classification.

B. Task 2: Relational Trigger Detection

We evaluate multi-label relational trigger detection using standard classification metrics. Table VI reports performance on the held-out test set.

TABLE VI
TASK 2 EVALUATION RESULTS (TRIGGER DETECTION).

Metric	Score
Macro Precision	0.358
Macro Recall	0.291
Macro F1	0.295
Micro Precision	0.641
Micro Recall	0.614
Micro F1	0.627

The disparity between macro- and micro-averaged scores reflects substantial label imbalance. Frequent relational stressors (e.g., communication-related issues) are captured more reliably, whereas low-frequency triggers such as timezone mismatch remain challenging.

C. Task 3: Temporal Phase Prediction

Temporal phase prediction is evaluated as a single-label classification task. Table VII summarizes test-set performance using accuracy and macro-F1.

D. Summary and Discussion

Across all tasks, the results demonstrate that:

TABLE VII
TASK 3 EVALUATION RESULTS (TEMPORAL PHASE CLASSIFICATION).

Metric	Score
Accuracy	0.704
Macro F1	0.518

- multi-label disorder classification is feasible, but clinically specific and lower-frequency disorders remain under-detected,
- relational triggers are harder to identify, partly due to subtle linguistic cues and label sparsity,
- temporal stress phase classification achieves the strongest macro-F1 performance among the three tasks.

Taken together, these results validate the difficulty and necessity of modelling psychological stress in long-distance relationships using naturalistic text.

VI. DISCUSSION

The results presented in Section VI underscore both the feasibility and the significant complexity of modeling relationally mediated psychiatric symptom categories. By anchoring NLP tasks to DSM-5-TR and ICD-11 criteria within the paradigmatic setting of Long-Distance Relationships (LDRs), this work provides empirical answers to the four research questions posed in Section I.

A. Clinical Validity and Complexity (RQ1 & RQ4)

Our evaluation of Task 1 (Multi-Label Disorder Classification) reveals that clinically grounded inference is substantially more challenging than standard depression detection. While RoBERTa-base achieves a Macro-F1 of 0.452 only 6.8 percentage points above the TF-IDF baseline (0.384) this performance lags significantly behind typical benchmarks for binary depression classification, where transformer models routinely exceed 0.85 F1 on datasets like RSDD and SMHD [10], [11].

This performance gap is primarily driven by the difficulty of modeling comorbidity and symptom overlap. Models struggle to distinguish between *Adjustment Disorder* and *Major Depressive Disorder (MDD)*. In clinical practice, this distinction relies on determining whether the symptoms are a proportionate response to a stressor. Current transformer architectures, which rely heavily on lexical co-occurrence, struggle with contextual reasoning about stressor-linked language. They successfully identify ”sadness” (high recall) but fail to discern its pathological distinctness (low precision for Adjustment Disorder). This suggests that future work must move beyond semantic embeddings toward neuro-symbolic or reasoning-based frameworks that can explicitly model the stressor-response relationship. The severe label imbalance, where Insomnia comprises only 1.2% of samples compared to 74.5% for Adjustment Disorder, partially explains the performance gap. However, even when controlling for class size via weighted loss functions (results not shown), MDD detection remained poor ($F1 < 0.1$). This suggests a more fundamental issue: **symptom ambiguity**. Posts expressing "I can't sleep because I miss them" could indicate either transient adjustment distress or clinical Insomnia Disorder, depending on duration and functional impairment criteria that are often implicit or absent in social media text.

B. The Diagnostic Value of Relational Triggers (RQ2)

A key contribution of this work is establishing that relational triggers are clinically important features for understanding psychiatric symptoms in digitally mediated relationships. While we did not conduct a formal ablation study in this iteration, the dataset structure and model performance provide suggestive evidence for their predictive value.

Observational Evidence from Performance Patterns: Separation Anxiety Disorder achieves achieves the strongest minority-class recall (0.72) compared to other clinically specific and lower-frequency disorders (MDD F1 = 0.00, Insomnia F1 = 0.00). This differential performance is not coincidental. SAD is fundamentally defined by excessive anxiety concerning separation from attachment figures. Posts labeled with SAD naturally contain explicit language around separation events, reunion dynamics, communication gaps, and abandonment fears precisely the relational triggers captured in our annotation schema. By contrast, MDD and Insomnia, while often triggered by relational stressors, are less directly defined by separation-specific language.

Implications for Future Work: This observational pattern motivates formal ablation studies as a key next step. Specifically, we propose masking relational trigger text (e.g., removing spans mentioning "silence gaps," "ghosting," "time zone mismatch," or "commitment uncertainty") and re-evaluating the model to quantify the performance drop. We hypothesize that such an ablation would reveal a 10-15% drop in macro-F1 for Separation Anxiety Disorder detection, validating that relational context is not merely helpful but necessary for accurate clinical inference in LDRs.

This finding answers RQ2: **computational models can and must utilize interactional triggers to accurately model symptom-aligned language in relational contexts**. In LDRs, a "silence gap" is not merely a communication log; it is a relational stressor that may intensify reported anxiety into active panic. This advocates for a shift in Digital Phenotyping: rather than analyzing user text in isolation, effective systems must analyze the **interactional metadata** (e.g., response latency, patterns of availability) alongside narrative text to achieve clinical validity.

Addressing the Trigger Distribution Puzzle: Interestingly, the trigger distribution reveals an unexpected finding: while "Timezone Mismatch" is frequently cited in LDR psychology literature, it appears in only 29 posts (1.6%). This discrepancy may reflect selection bias users experiencing circadian stress may normalize it rather than framing it as a "problem" worthy of posting. Conversely, "Commitment Ambiguity" dominates at 66.2% of posts, suggesting that *meta-relational uncertainty* (e.g., "Are we really together?") is a more salient stressor in Reddit communities than logistical challenges like time zones. This raises an important methodological question: **does RSPC capture the full spectrum of LDR stressors, or does the Reddit community sample bias toward emotional/psychological triggers over logistical ones?** Future work incorporating diverse platforms and populations would address this.

The centrality of relational triggers in RSPC's design positions this dataset uniquely to evaluate whether incorporating interactional context improves computational mental health models a research direction largely absent from existing mental health NLP benchmarks.

C. Temporal Dynamics and the "LDR Cycle" (RQ3)

The results from Task 3 (Temporal Stress Forecasting) confirm that mental health in LDRs is cyclical rather than linear.

This confirms RQ3: incorporating the temporal context of the "LDR Cycle" (Pre-Separation $\rightarrow$ Separation $\rightarrow$ Reunion) is essential for accurate forecasting. A user expressing distress in the *Pre-Separation* phase (anticipatory grief) has a fundamentally different risk trajectory than a user expressing distress in the *Post-Reunion* phase (post-event blues). Existing "early detection" systems often model risk as a monotonic increase; our findings suggest that for relational disorders, intervention systems must be "phase-aware," delivering different coping strategies depending on where the user sits in the separation cycle.

D. Generalizability to Broader Digital Contexts

While grounded in LDRs, the **RSPC** dataset models a universal dynamic: *technologically mediated separation*. These findings suggest that similar stressor–symptom associations may be observed in other digitally mediated contexts, such as:

- **Remote Caregiving:** Where family members monitor elderly relatives via digital tools, often leading to "remote caregiver burnout."
- **Migration and Displacement:** Where refugees maintain family bonds solely through apps like WhatsApp, creating high-stakes "digital togetherness."
- **Pandemic Isolation:** The COVID-19 lockdowns replicated the LDR experience on a global scale, making our findings on "circadian desynchronization" and "video-call fatigue" universally relevant.

VII. Limitations and Future Work

While RSPC introduces a novel relationally grounded benchmark for computational mental-health research, several limitations remain.

First, the dataset exhibits substantial class imbalance across both psychiatric symptom categories and relational stressors. Rare labels such as MDD, INSOMNIA, and stressors like TIMEZONE MISMATCH and SOCIAL MEDIA SURVEILLANCE appear infrequently, leading to near-zero F_1 scores. This highlights the difficulty of learning robust decision boundaries for clinically specific and minority phenomena using standard supervised objectives.

Second, although our models incorporate relational context implicitly through text, they do not explicitly model interaction structure (e.g., message sequences, bidirectionality, or partner perspectives). As a result, relational stress is still inferred from individual posts rather than from dyadic or longitudinal interaction data.

Third, annotations are derived from Reddit self-disclosures, which may not generalize to other populations or clinical settings. Self-reported distress can differ from clinically assessed diagnoses, and platform-specific norms may influence language use.

Future work can address these limitations in several ways. Data augmentation or cost-sensitive learning may improve performance on underrepresented labels. Incorporating temporal modeling and conversational structure could enable more faithful representations of relational dynamics. Extending RSPC with multimodal signals (e.g., timestamps, communication gaps, or interaction networks) and validating models against clinician-reviewed data would further strengthen the clinical relevance and generalizability of this line of research.

VIII. Conclusion

This paper introduced the **Relational Stress and Psychiatry Corpus (RSPC)**, the first benchmark dataset to provide an NLP benchmark for studying how relational stressors co-occur with clinically-grounded psychiatric categories. By moving beyond static, individual-level distress detection, we established a new paradigm for relationship-aware computational psychiatry.

Our experiments demonstrated that while transformer-based models can detect broad symptom patterns, they struggle with the fine-grained, context-based reasoning required to distinguish co-morbid conditions like Adjustment Disorder and Separation Anxiety Disorder. We provide empirical evidence that relational triggers are strongly associated with improved detection of separation-related disorders (e.g., silence gaps, circadian misalignment) and temporal phase information significantly improves predictive performance, highlighting that contextual information can be valuable when modelling mental-health–related language.

As digital platforms increasingly mediate human intimacy from Long-Distance Relationships to remote caregiving and migration-induced separation the need for context-sensitive AI systems becomes paramount. The RSPC dataset provides the necessary foundation for developing the next generation of digital mental health tools: systems that are not only clinically accurate but also deeply attuned to the relational rhythms of users' lives. Future work will focus on integrating multi-modal signals and expanding this framework to cross-cultural settings to further bridge the gap between affective computing and clinical reality. We emphasize that RSPC models predict symptom-category labels derived from user-generated language and do not constitute clinical diagnosis.

References

[1] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "Clpsych 2015 shared task: Depression and ptsd on twitter," in *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015, pp. 31–39.

[2] D. E. Losada, F. Crestani, and J. Parapar, "Overview of erisk: early risk prediction on the internet," in *International conference of the cross-language evaluation forum for european languages*. Springer, 2018, pp. 343–361.

[3] S. Chancellor, M. L. Birnbaum, E. D. Caine, V. M. Silenzio, and M. De Choudhury, "A taxonomy of ethical tensions in inferring mental health states from social media," in *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 79–88.

[4] K. Harrigian, C. Aguirre, and M. Dredze, "On the state of social media data for mental health research," in *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, 2021, pp. 15–24.

[5] D. N. Milne, G. Pink, B. Hachey, and R. A. Calvo, "Clpsych 2016 shared task: Triaging content in online peer-support forums," in *Proceedings of the third workshop on computational linguistics and clinical psychology*, 2016, pp. 118–127.

[6] S. MacAvaney, A. Mittu, G. Coppersmith, J. Leintz, and P. Resnik, "Community-level research on suicidality prediction in a secure environment: Overview of the clpsych 2021 shared task," in *Proceedings of the seventh workshop on computational linguistics and clinical psychology: Improving access*, 2021, pp. 70–80.

[7] D. E. Losada, F. Crestani, and J. Parapar, "Overview of erisk at clef 2019: Early risk prediction on the internet (extended overview)." *CLEF (Working Notes)*, vol. 4, p. 21, 2019.

[8] ——, "erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations," in *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 2017, pp. 346–360.

[9] J. Parapar, A. Perez, X. Wang, and F. Crestani, "Overview of erisk 2025: Early risk prediction on the internet," in *International conference of the cross-language evaluation forum for European languages*. Springer, 2025, pp. 242–265.

[10] A. Yates, A. Cohan, and N. Goharian, "Depression and self-harm risk assessment in online forums," *arXiv preprint arXiv:1709.01848*, 2017.

[11] A. Cohan, B. Desmet, A. Yates, L. Soldaini, S. MacAvaney, and N. Goharian, "Smhd: a large-scale resource for exploring online language usage for multiple mental health conditions," *arXiv preprint arXiv:1806.05258*, 2018.

[12] E. Turcan and K. McKeown, "Dreaddit: A reddit dataset for stress analysis in social media," *arXiv preprint arXiv:1911.00133*, 2019.

[13] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik, "Expert, crowdsourced, and machine assessment of suicide risk via online postings," in *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, 2018, pp. 25–36.

[14] S. Guarino, A. Scarlata, and A. Gigli, "A characterization of online suicide risk and mental health crises," in *Proceedings of the 2022 IEEE International Conference on Digital Health (ICDH)*, 2022, pp. 11–20.

[15] S. Zanwar and J. Daniel, "Cross-domain modeling of mental health comorbidities from social media," in *Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022, pp. 342–349.

[16] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: A critical review," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1–11, 2020.

[17] S. MacAvaney, A. Cohan, and N. Goharian, "Capturing the temporal dynamics of depressive language in social media," in *EMNLP 2018*, 2018, pp. 2968–2978.

[18] A. Tsakalidis, M. Liakata, T. Damoulas, and A. I. Cristea, "Now or later? predicting and maximising success in psychotherapy using natural language processing," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, 2018, pp. 393–403.

[19] R. Sawhney, H. Joshi, S. Gandhi, and R. R. Shah, "A time-aware transformer based model for suicide ideation detection on social media," in *EMNLP 2020*, 2020, pp. 7685–7697.

[20] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez, "A text classification framework for simple and effective early depression detection over social media streams," *Expert Systems with Applications*, vol. 133, pp. 182–197, 2019.

[21] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing neural networks and dictionary-based features for early depression detection," in *CLEF 2018 Working Notes*, 2018.

[22] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning for suicide ideation detection: A review," in *IEEE Reviews in Biomedical Engineering*, 2020.

[23] R. Mishra and A. Varshney, "Context-aware hierarchical attention network for depression detection," in *Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM)*, 2019.

[24] V. R. K. Garimella, I. Weber, and M. De Choudhury, "Love, heartbreak, and twitter: Measuring the scale and effect of romantic relationships," in *CHI 2014*, 2014, pp. 1–10.

[25] E. Adar, L. A. Adamic, L. Zhang, and R. M. Lukose, "Implicit structure and the dynamics of blogspace," *Workshop on the Weblogging Ecosystem*, 2004.

[26] L. Backstrom and J. Kleinberg, "Romantic partnerships and the dispersion of social ties," in *CSCW 2014*, 2014, pp. 831–841.

[27] L. C. Jiang and J. T. Hancock, "Computer-mediated communication, conflict, and intimacy in long-distance romantic relationships," *Annals of the International Communication Association*, vol. 37, no. 1, pp. 343–373, 2013.

[28] L. Stafford, *Maintaining Long-Distance and Cross-Residential Relationships*. Lawrence Erlbaum Associates, 2006.

[29] C. Licoppe, "'connected' presence: The emergence of a new repertoire for managing social relationships in a changing communication technoscape," *Environment and Planning D: Society and Space*, vol. 22, no. 1, pp. 135–156, 2004.

[30] C. Neustaedter and S. Greenberg, "The benefits and challenges of video calling for long-distance relationship maintenance," in *CHI 2015*, 2015, pp. 1–10.

[31] L. Lucero and S. Jones, "Absence makes the heart grow fonder: Effects of geographical separation on intimacy in romantic relationships," in *CSCW 2014*, 2014.

[32] S. T. Tong and J. B. Walther, "Computer-mediated communication and romance," *Handbook of Interpersonal Communication*, pp. 61–86, 2004.

[33] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*. Washington, DC: American Psychiatric Association Publishing, 2022.

[34] World Health Organization, *International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision)*. World Health Organization, 2019.

[35] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O'Dea, C. Paris, and H. Christensen, "We feel: Mapping emotion on twitter," in *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, 2015, pp. 1246–1252.

[36] M. K. Shear, N. Simon, M. Wall, S. Zisook, R. Neimeyer, N. Duan, C. Reynolds, B. Lebowitz, S. Sung, A. Ghesquiere *et al.*, "Complicated grief and related bereavement issues for dsm-5," *Depression and Anxiety*, vol. 28, no. 2, pp. 103–117, 2011.

[37] B. Milrod, J. C. Markowitz, and A. Mvartin, "Separation anxiety disorder in school-age children: What health care providers should know," *Journal of Clinical Psychiatry*, 2014.

[38] A. Benton, G. Coppersmith, and M. Dredze, "Ethical research protocols for social media health research," in *Proceedings of the 1st ACL Workshop on Ethics in Natural Language Processing*, 2017, pp. 94–102.

[39] M. Conway, "Ethical issues in using twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature," *Journal of medical Internet research*, vol. 16, no. 12, p. e290, 2014.

[40] S. Chancellor and M. De Choudhury, "Norms matter: Contrasting social support around behavior change in online health communities," *CHI 2019*, 2019.

[41] L. Stark, "Facial recognition, emotion, and the politics of predictive policing," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2019.

[42] H. McKelvey, R. Brady, and P. Scott, "Clinical early warning systems: An ethical analysis," *Journal of Medical Ethics*, 2019.

[43] N. Martinez-Martin, "Privacy-preserving mental health monitoring wearables," *IEEE Technology and Society Magazine*, 2019.