

# KCL-Cogstack at PsyDefDetect: A Hierarchical Approach to Detecting Defense Mechanisms in Supportive Dialogue

Shubham Agarwal<sup>1</sup>, Thomas Searle<sup>1,3</sup>, Richard Dobson<sup>1,2,3</sup>

<sup>1</sup>Department of Biostatistics & Health Informatics, King’s College London, London, U.K.

<sup>2</sup>Institute of Health Informatics, University College London, London, U.K.

<sup>3</sup>CogStack Limited, London, U.K.

Correspondence: [shubham.agarwal@kcl.ac.uk](mailto:shubham.agarwal@kcl.ac.uk)

## Abstract

We present our system for the PsyDefDetect shared task, which focuses on detecting and classifying psychological defense mechanisms in peer emotional support conversations. Our core contribution is a hierarchical classification framework that structures prediction as a coarse-to-fine pipeline over a clinically validated label hierarchy, grounded in the Defense Mechanism Rating Scales (DMRS). Through systematic experimentation with flat fine-tuning, few-shot prompting, and hierarchical classification, we demonstrate that explicitly modelling the structured relationships among defense levels offers a more effective alternative to flat classification, achieving a macro F1 of 0.23 on the official test set.

## 1 Introduction

As digital and transcription-based health records capture increasing volumes of patient–clinician interactions, there is a growing opportunity to analyze how mental states and psychological processes manifest in language (Zimmerman et al., 2022; Drew et al., 2001). Transcripts of therapy sessions, clinical interviews, and patient communications provide rich evidence of cognitive and emotional patterns that can inform mental health assessment and care (Ewbank et al., 2021). This offers substantial opportunities to better understand patient experiences and mental states, and to facilitate timely, accurate, and psychologically informed interpretation of patient communications — essential for ensuring that individuals feel heard, supported, and engaged in their care (Drew et al., 2001). Such data can further support a wide range of mental health applications, from screening and monitoring to personalized assistance (Luxton, 2014).

Among the many phenomena expressed in clinical dialogue, psychological defense mechanisms are of particular importance. Psychological defense mechanisms are unconscious or automatic

strategies that individuals use to manage distress, reduce emotional conflict, and protect themselves from anxiety or perceived threats (Freud, 1936; Vaillant, 1992). Because these mechanisms are often expressed indirectly in language, they provide a clinically meaningful target for automatic analysis of patient text and dialogue (Cramer, 2000).

Recent work on automatically identifying defense mechanisms from text has begun to show that language models can capture clinically meaningful defensive cues in dialogue. A notable study applied RoBERTa-based models (Liu et al., 2019) to interviewer–participant talk–turn pairs from adult attachment interview transcripts and demonstrated that language models can distinguish the presence of defenses and detect several common mechanisms, including repression, intellectualization, reaction formation, and undoing (Tasca et al., 2023). This provides an important proof of concept that defensive functioning can be modeled computationally from conversational data.

The broader psychodynamic literature offers a strong theoretical foundation for this task. The Defense Mechanism Rating Scales (DMRS) (Di Giuseppe and Perry, 2021) are a clinically validated, hierarchically organised system for rating defensive maturity in clinical transcripts, ranging from immature action-based defenses to high-adaptive coping strategies. Originally developed for use by trained clinicians, the DMRS have long been used to assess defensive functioning, and recent work continues to support their relevance in contemporary assessment and computational annotation settings.

However, despite these advances, automatic recognition of defense mechanisms from text remains an underexplored problem, particularly in the setting of supportive conversations and multi-turn dialogue (Na et al., 2025). The PsyDefDetect shared task (Na et al., 2026a) addresses this gap by introducing a benchmark for detecting seven de-

fense mechanisms plus two auxiliary labels in text, thereby providing a clinically grounded testbed for future NLP methods.

## 2 Background

### 2.1 PsyDefDetect shared task

The *PsyDefDetect* shared task (Na et al., 2026a) introduces a benchmark for the detection and classification of psychological defense mechanisms in peer emotional support conversations. The task is motivated by recent work on computational modeling of defensive functioning in dialogue, which shows that transformer-based language models can distinguish the presence of defenses and identify common mechanisms from conversational data. *PsyDefDetect* extends this line of research by providing an annotated dataset of conversations explicitly labeled for defense levels, linked to a clinically grounded taxonomy based on the DMRS framework.

### 2.2 Dataset and setup

The dataset, PSYDEFCONV (Na et al., 2026b), is derived from a stratified subset of the ESConv corpus (Liu et al., 2021) and contains 200 dialogues and 4709 utterances, of which 2336 are help-seeker turns annotated for defensive functioning. Each instance consists of a multi-turn dialogue and a target utterance; systems must classify the target into one of nine categories: seven hierarchical DMRS defense levels, a *No Defenses* label for phatic or functional turns, and a *Needs More Information* label for utterances lacking sufficient context. By providing a structured, publicly available benchmark, *PsyDefDetect* aims to stimulate work on interpretable and clinically meaningful methods for understanding how individuals express psychological distress in language.

## 3 Methodology

We explore several modeling strategies for detecting psychological defense mechanisms from supportive conversations. First, we investigate standard supervised learning, where a BERT model (Devlin et al., 2019) is fine-tuned to classify text directly into the nine defense-mechanism classes. This *flat* classification approach treats all mechanisms as independent labels and does not explicitly model higher-level groupings or developmental levels.

Second, we consider few-shot prompting of Large Language Models (LLMs), where the task is cast as in-context learning rather than fine-tuning. This approach avoids the need for large annotated training sets and can capture nuanced patterns via the model’s pre-existing knowledge of language and psychology.

Finally, we propose a hierarchical classification framework grounded in clustering. Rather than treating the nine defense mechanisms as independent labels, we first group them into higher-level clusters based on semantic and developmental similarity. These clusters then structure a hierarchical classifier that first predicts the level or cluster of defensive functioning before refining the prediction to the specific mechanism. This approach reflects the theoretical structure of the DMRS taxonomy and enables more structured and interpretable modeling of the task.

Our experimentation utilized a shared university resource machine with 2 Nvidia A100 GPUs via KCL CREATE (King’s College London e-Research team, 2025).

### 3.1 Approaches

#### 3.1.1 Fine-tuning BERT

We fine-tune a BERT-based encoder for defense-mechanism classification. Specifically, we use `bert-base-uncased` as the backbone model and apply Low-Rank Adaptation (LoRA) (Hu et al., 2022) to efficiently adapt the encoder to the target task. Given the limited size of the annotated dataset and the clinically nuanced nature of the labels, parameter-efficient fine-tuning is a natural choice, as LoRA introduces a small number of trainable parameters while keeping the original model weights largely frozen, reducing the risk of overfitting. A key challenge is that full dialogue context often exceeds BERT’s 512-token limit and also contains noise and less relevant turns. We therefore explore two context representation strategies:

- *last-3 turns*, where only the three most recent dialogue turns along with the target utterance are included as context
- *summarised context*, where all preceding turns except the final one are condensed into a summary using an LLM (Gemma 4B (Team, 2025)\*), which is then concatenated with the final

\*<https://huggingface.co/google/gemma-3-4b-it>

turn and the target utterance

Both representations are kept within the 512-token limit. For training, we use the Adam optimizer with a learning rate of  $1e-4$ , a dropout rate of 0.25, and gradient clipping with a maximum norm of 1.0, with a batch size of 64. The model is trained on an 80–20 train–validation split of the official training data.

### 3.1.2 Prompting LLM

We leverage in-context learning via few-shot prompting, providing a structured prompt that describes the task, introduces the nine labels with brief definitions for each class, and includes a small set of example instances, each containing a dialogue segment and its corresponding ground-truth label. We implement this using MedGemma-4B (Sellersgren et al., 2025), a model optimized for clinical and health-related language, which we hypothesize may better capture the nuanced psychological language present in emotional support conversations. Prompt design was iteratively refined based on empirical performance. Full prompt templates are provided in Appendix A.4.

### 3.1.3 Hierarchical classification

In addition to flat fine-tuning and few-shot prompting, we explore a hierarchical classification strategy designed to reflect the structured nature of the defense-mechanism space. Rather than treating all nine labels as independent classes, we first group them into higher-level clusters before performing final classification.

To construct these groupings, we apply spectral clustering (Von Luxburg, 2007) over sentence embeddings of the training utterances using an RBF kernel over the normalised graph Laplacian to obtain label-level clusters. The resulting groupings were validated by a psychiatrist at South London and Maudsley NHS Foundation Trust (SLaM) for clinical plausibility.

The resulting hierarchy, summarized in Appendix A.2 (Table 2), organizes the nine labels into three levels of increasing specificity. This coarse-to-fine pipeline exploits both the natural hierarchy of defensive functioning and the clinically meaningful relationships among the labels, providing a more structured alternative to flat classification. Here as well we use bert-base-uncased as the classification model. The pipeline consists of a separate BERT models trained for each node in the hierarchy, with predictions at each level routing the

input to the appropriate child classifier. Errors at higher levels propagate to subsequent stages, as no recovery mechanism is employed.

## 3.2 Evaluation

We evaluate all systems using F1-score as the primary metric, following the official PsyDefDetect evaluation protocol. We additionally report accuracy and weighted F1-score to provide a more complete picture of model behaviour across the nine categories. We report results on both the validation set (20% of the dev set), and the official test set evaluated through the CodaBench platform. Validation results are used to guide model selection and hyperparameter tuning, while test set results reflect final system performance under the shared task evaluation conditions.

## 4 Results

### 4.1 Fine-tuned BERT

The fine-tuned BERT model achieves moderate performance on both the validation and test sets. On the validation split, it obtains a macro F1 of 0.19. On the official test set, performance drops to a macro F1 of 0.15, suggesting some degree of overfitting to the development split. These results suggest that while standard supervised fine-tuning provides a reasonable starting point, it struggles to fully capture the nuanced distinctions between defense mechanisms.

### 4.2 LLM Prompting

The prompting-based approach performs worse than fine-tuned BERT on both splits. On the entire development set (no training required), it reaches a macro F1 of 0.12, and on the official test set, performance is comparable at 0.13. This indicates that few-shot prompting alone is not sufficient for reliably identifying defense mechanisms in this task, likely because the label space is clinically subtle and the examples are highly heterogeneous.

### 4.3 Hierarchical BERT

Hierarchical BERT is the strongest-performing approach across both validation and test evaluations. On the validation split, it achieves a macro F1 of 0.24, achieving the rank of 17/21 registered teams, outperforming both the fine-tuned BERT and LLM prompting approaches. On the test set, it again ranks highest with an F1 of 0.23. These results suggest that explicitly modelling the structure among

Table 1: Overall performance across approaches on the validation and test sets

Model	<i>Micro</i>			<i>Macro</i>			<i>Weighted</i>		
	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	F1
<i>Validation set performance</i>									
Fine-tuned BERT	0.38	0.38	0.38	0.18	0.19	0.19	0.33	0.38	0.35
LLM *	0.31	0.31	0.31	0.14	0.13	0.12	0.31	0.35	0.32
<b>Hierarchical BERT</b>	<b>0.47</b>	<b>0.48</b>	<b>0.47</b>	<b>0.26</b>	<b>0.24</b>	<b>0.24</b>	<b>0.48</b>	<b>0.61</b>	<b>0.52</b>
<i>Test set performance</i>									
Model	Recall			Precision			F1		
Fine-tuned BERT	0.17			0.26			0.15		
LLM	0.16			0.17			0.13		
<b>Hierarchical BERT</b>	<b>0.25</b>			<b>0.29</b>			<b>0.23</b>		

\* Evaluation results on entire dev set

defense mechanisms is beneficial and that hierarchical classification offers a more effective way to represent the label space than flat prediction.

Level-wise evaluation, provided in Appendix A.3 (Table 3), reveals that the hierarchy performs strongly at coarse discrimination but degrades at finer granularities, suggesting that distinguishing between specific mechanisms within a cluster remains the key challenge. Additionally, the confusion matrices for all approaches are also provided in Appendix A.1.

## 5 Discussion

Our approach to the PsyDefDetect shared task evolved through several iterations, each motivated by the limitations of the previous. We began with a flat fine-tuned BERT classifier, which provided a reasonable starting point but exposed a key weakness: the model heavily biased toward frequent classes, particularly *No Defenses* and *High-Adaptive Defenses*, while failing entirely to predict the *Needs More Information* label. This suggests that standard supervised fine-tuning struggles to handle the class imbalance inherent in the PSYDEFCONV dataset.

We then evaluated a few-shot prompting approach using MedGemma-4B, and while this avoided the need for fine-tuning, it proved even more prone to prediction collapse, effectively reducing the label space to a handful of classes and ignoring clinically distinct mechanisms such as *Minor Image-Distorting* and *Neurotic* defenses entirely. This indicates that few-shot prompting alone is insufficient for a task requiring fine-grained,

context-dependent clinical judgment.

These observations motivated our hierarchical classification framework, which explicitly models the structured relationships among defense levels rather than treating all labels as independent. By introducing coarse-to-fine prediction grounded in a clinically validated label hierarchy, this approach achieved the strongest performance across both development and test evaluations. Notably, the hierarchical approach produced more balanced predictions across the label space compared to flat classification, though rare and clinically subtle classes remained challenging across all systems.

Compared to the best results reported in the dataset paper (Na et al., 2026b), where a macro F1 of 0.31 is achieved, our results highlight the difficulty of this task and suggest that further gains may require richer contextual representations, larger models, or more sophisticated handling of class imbalance.

## 6 Conclusion & Future Work

This paper presents a hierarchical classification framework for detecting psychological defense mechanisms in the PsyDefDetect shared task, incorporating LLM-based dialogue summarisation to handle long conversational contexts.

By structuring prediction as a coarse-to-fine pipeline over a psychiatrist-validated label hierarchy, our approach achieves a F1 of 0.23 on the official PsyDefDetect test set, outperforming both flat fine-tuning and few-shot prompting baselines. The results demonstrate that explicitly modelling the structured relationships among defense levels

offers a more effective and interpretable alternative to flat classification for this clinically nuanced task.

Our future work will focus on addressing class imbalance and exploring domain-adapted mental health language models to better capture subtle defensive cues.

## Limitations

Our work presents a hierarchical classification framework for the PsyDefDetect shared task, evaluated on both a development split and the official unseen test set. However, the proposed system is limited in a number of ways.

First, all three approaches struggle with class imbalance in the PSYDEFCONV dataset, where High-Adaptive Defenses account for over 50% of annotated turns. This leads to prediction bias toward frequent classes and poor performance on rare but clinically important categories such as Major Image-Distorting and Neurotic defenses.

Second, the dataset is relatively small, comprising 200 dialogues drawn from a single corpus of peer emotional support conversations. This limits the generalisability of our findings to other clinical settings, such as formal psychotherapy or psychiatric interviews, where defensive functioning may manifest differently.

Third, the hierarchical pipeline chains independent classifiers with hard decisions at each level, meaning that misclassifications at coarser levels propagate to finer levels with no recovery mechanism. Future work could explore soft or joint training approaches to mitigate this.

Finally, deployment of such systems in real clinical environments would require careful attention to data privacy, regulatory compliance (e.g. GDPR), and the ethical implications of automated psychological assessment, particularly given the sensitivity of mental health data.

## Acknowledgments

This work was supported by Health Data Research UK, an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities. SA, TS, RD are part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. RD is also supported by The National Institute for Health

Research University College London Hospitals Biomedical Research Centre. This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

- Phebe Cramer. 2000. Defense mechanisms in psychology today: Further processes for adaptation. *American psychologist*, 55(6):637.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.
- Mariagrazia Di Giuseppe and J Christopher Perry. 2021. The hierarchy of defense mechanisms: Assessing defensive functioning with the defense mechanisms rating scales q-sort. *Frontiers in psychology*, 12:718440.
- Paul Drew, John Chatwin, and Sarah Collins. 2001. Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expectations*, 4(1):58–70.
- MP Ewbank, R Cummins, V Tablan, A Catarino, S Buchholz, and AD Blackwell. 2021. Understanding the relationship between patient language and outcomes in internet-enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts. *Psychotherapy Research*, 31(3):300–312.
- Sigmund Freud. 1936. Inhibitions, symptoms and anxiety. *The Psychoanalytic Quarterly*, 5(1):1–28.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, and 1 others. 2022. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3.
- King's College London e-Research team. 2025. [King's computational research, engineering and technology environment \(CREATE\)](#).
- Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. In *Proceedings of the 59th annual meeting*

- of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)*, pages 3469–3483.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- David D Luxton. 2014. Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice*, 45(5):332.
- Hongbin Na, Yining Hua, Zimu Wang, Tao Shen, Beibei Yu, Lilin Wang, Wei Wang, John Torous, and Ling Chen. 2025. [A survey of large language models in psychotherapy: Current landscape and future directions](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 7362–7376, Vienna, Austria. Association for Computational Linguistics.
- Hongbin Na, Zimu Wang, Zhaoming Chen, Yining Hua, Rena Gao, Kailai Yang, Ling Chen, Wei Wang, Shaoxiong Ji, John Torous, and Sophia Ananiadou. 2026a. Overview of the psydefdetect shared task at bionlp 2026: Detecting levels of psychological defense mechanisms in supportive conversations. In *Proceedings of the 25th Workshop on Biomedical Language Processing*, San Diego, USA. Association for Computational Linguistics.
- Hongbin Na, Zimu Wang, Zhaoming Chen, Peilin Zhou, Yining Hua, Grace Ziqi Zhou, Haiyang Zhang, Tao Shen, Wei Wang, John Torous, Shaoxiong Ji, and Ling Chen. 2026b. You never know a person, you only know their defenses: Detecting levels of psychological defense mechanisms in supportive conversations. In *Findings of the Association for Computational Linguistics: ACL 2026*, San Diego, USA. Association for Computational Linguistics.
- Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cian Hughes, Charles Lau, and 1 others. 2025. Medgemma technical report. *arXiv preprint arXiv:2507.05201*.
- Anthony N Tasca, Samantha Carlucci, James C Wiley, Matthew Holden, Ahmed El-Roby, and Giorgio A Tasca. 2023. Detecting defense mechanisms from adult attachment interview (aai) transcripts using machine learning. *Psychotherapy Research*, 33(6):757–767.
- Gemma Team. 2025. [Gemma 3](#).
- George E Vaillant. 1992. *Ego mechanisms of defense: a guide for clinicians and researchers*. American Psychiatric Pub.
- Ulrike Von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416.
- Marc Zimmerman and 1 others. 2022. [Linguistic measures of psychological distance track symptom levels and treatment progress in psychotherapy](#). *Proceedings of the National Academy of Sciences*.

## A Appendix

### A.1 Confusion Matrices

#### A.1.1 Fine-tuned BERT

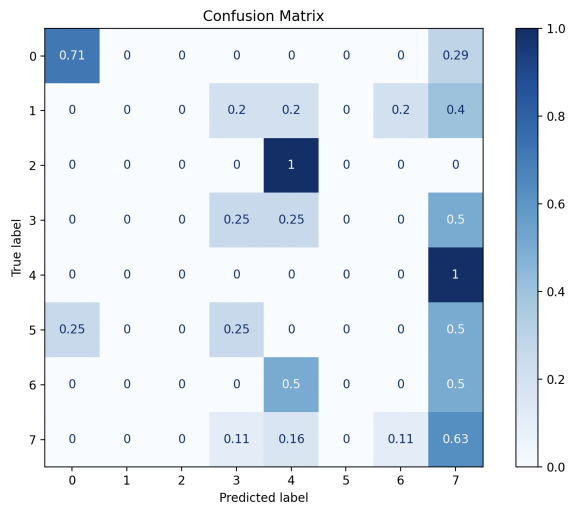


Figure 1: Confusion matrix for fine-tuned BERT

\*Since class 8 wasn't predicted, it's not shown in the confusion matrix

#### A.1.2 LLM

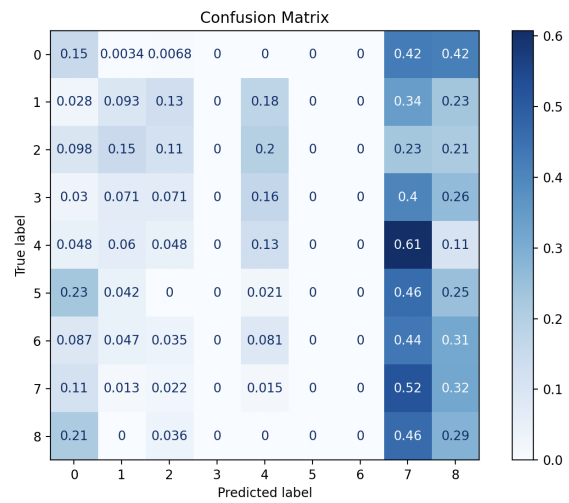


Figure 2: Confusion matrix for LLM

#### A.1.3 Hierarchical classifier

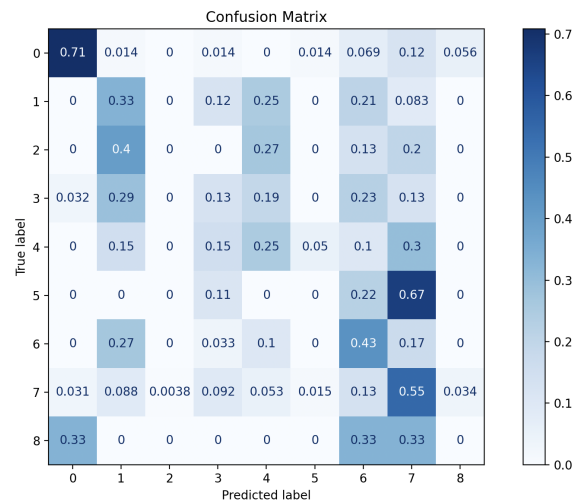


Figure 3: Confusion matrix for hierarchical classifier

## A.2 Hierarchical label clustering

Table 2: Hierarchical label clustering used in the three-level classification pipeline

Level	Cluster	Classes
Level 1	Auxiliary	No Defenses, Needs More Info
	High-Adaptive	High-Adaptive Defenses
	Maladaptive	Action, Major Image Distorting, Disavowal, Minor Image Distorting, Neurotic, Obsessional
Level 2	Auxiliary → N/A (terminal)	No Defenses, Needs More Info
	Maladaptive → Externalizing / Image-Distorting	Action, Major Image Distorting, Minor Image Distorting
	Maladaptive → Avoidant / Internal-Regulating	Disavowal, Neurotic, Obsessional
Level 3	Externalizing / Image-Distorting → N/A (terminal)	Action, Major Image Distorting, Minor Image Distorting
	Avoidant / Internal-Regulating → N/A (terminal)	Disavowal, Neurotic, Obsessional

## A.3 Hierarchical classification level-wise results

Table 3: Level-wise results for hierarchical classifier

Level	Recall	Precision	F1
<i>Validation set performance</i>			
Level 1	0.56	0.6	0.56
Level 2 - Auxiliary	0.71	0.97	0.79
Level 2 - Maladaptive	0.71	0.7	0.71
Level 3 - Externalizing	0.56	0.6	0.57
Level 3 - Avoidant	0.53	0.53	0.5

## A.4 Prompt for LLM

```

"""
<bos><start_of_turn>user You are a emotional conversation expert. Your task is to classify the given
conversation into one of the below given classes.

---
# Instructions:
1) The input will include a transcript of what the seeker is saying to the supporter.
2) Classify these into the given classes by taking into account information from the text, you
will have to infer it, the information won't be directly present.
3) Give reasoning on why you have chosen the class, explaining why it makes sense and pointing
out key evidence.
4) The output must be 'label' and 'reasoning' as the two keys in a json format.
5) Respond only with the label and reasoning, do not include additional commentary.

---
# Classes:
0: No Defenses; Mechanism: N/A; Definition: Functional utterances that maintain conversational
flow without engaging conflict.
1: Action Defenses; Mechanism: Passive Aggression, Help-Rejecting Complaining, Acting Out;
Definition: Distress is released by acting on the environment instead of reflecting.
2: Major Image-Distorting; Mechanism: Splitting (self/other), Projective Identification;
Definition: Reduces anxiety via all-good/all-bad distortions of self or other.

```

- 3: Disavowal Defenses; Mechanism: Denial, Rationalization, Projection, Autistic Fantasy;  
Definition: Rejects threatening reality by denying, excusing, blaming, or fantasizing.
- 4: Minor Image-Distorting; Mechanism: Devaluation/Idealization (self or other), Omnipotence;  
Definition: Softer distortions temporarily inflate or deflate self-esteem.
- 5: Neurotic Defenses; Mechanism: Repression, Dissociation, Reaction Formation, Displacement;  
Definition: Keeps unacceptable motives out of awareness; feelings surface indirectly.
- 6: Obsessional Defenses; Mechanism: Isolation of Affect, Intellectualization, Undoing; Definition:  
Uses excessive logic or symbolic acts to separate feelings from events.
- 7: High-Adaptive Defenses; Mechanism: Affiliation, Altruism, Anticipation, Humor, Self-Assertion,  
Self-Observation, Sublimation, Suppression; Definition: Mature coping that integrates  
emotion and thought to channel affect constructively.
- 8: Needs More Information; Mechanism: N/A; Definition: Label used when an utterance is too  
ambiguous or lacks context.

---

Question: Classify this text for me?

Input: {text}

<end\_of\_turn> <start\_of\_turn>model

"""