University of Milano-Bicocca
Department of Informatics, Systems,
and Communication (DISCo)
Master's Degree Program in Data Science
Academic Year 2023-2024

Large Language Models for Mental Health Support

Supervisor: Prof. Marco Viviani
Co-supervisor: Dott. Alessandro Raganato

Sabino Giuseppe 852287

Introduction

- Context and state of the art
- Dataset and technology
- Data Translation
- Experimental evaluation
- Conclusion

Context

- The growing efficiency of Large Language Models (LLMs) has facilitated their integration across various fields.
- Existing research has demonstrated their potential in the **mental health domain**, but most studies have been conducted predominantly in **English**.
- This research aims to extend the application of LLMs to the **Italian language** by translating three key datasets (**Dreaddit**, **DepSeverity**, and **SDCNL**) into Italian.
- **BERT, RoBERTa, Flan-T5, Qwen2 0.5B,** and **Qwen2 1.5B**, were trained and evaluated in italian in this context.
- This work represents a step towards improving mental health assessments in non-English contexts.

State of the art

MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare

- MentalBERT
- MentalRoBERTa

Mentalllm: Leveraging large language models for mental health prediction via online text data

- Mental-Alpaca
- Mental-Flan-T5
- CrossLanguage, Zero-shot prompting, Few-shot prompting with Alpaca, Alpaca-LoRA, FLAN-T5, GPT-3.5, and GPT-4

Contribution

- Performance analysis of various Large Language Models (LLMs) in the mental health domain, comparing results between the Italian and English languages.
- Showed that fine-tuning yields better performance compared to zero-shot predictions, even when using larger models.
- **Three Italian datasets** relevant to mental health research have been created, specifically translated versions of Dreaddit, DepSeverity, and SDCNL.
- **Four** new **Italian LLMs** have been trained, capable of recognizing stress, anxiety, depression, and suicidal intentions in social media posts.

Dataset

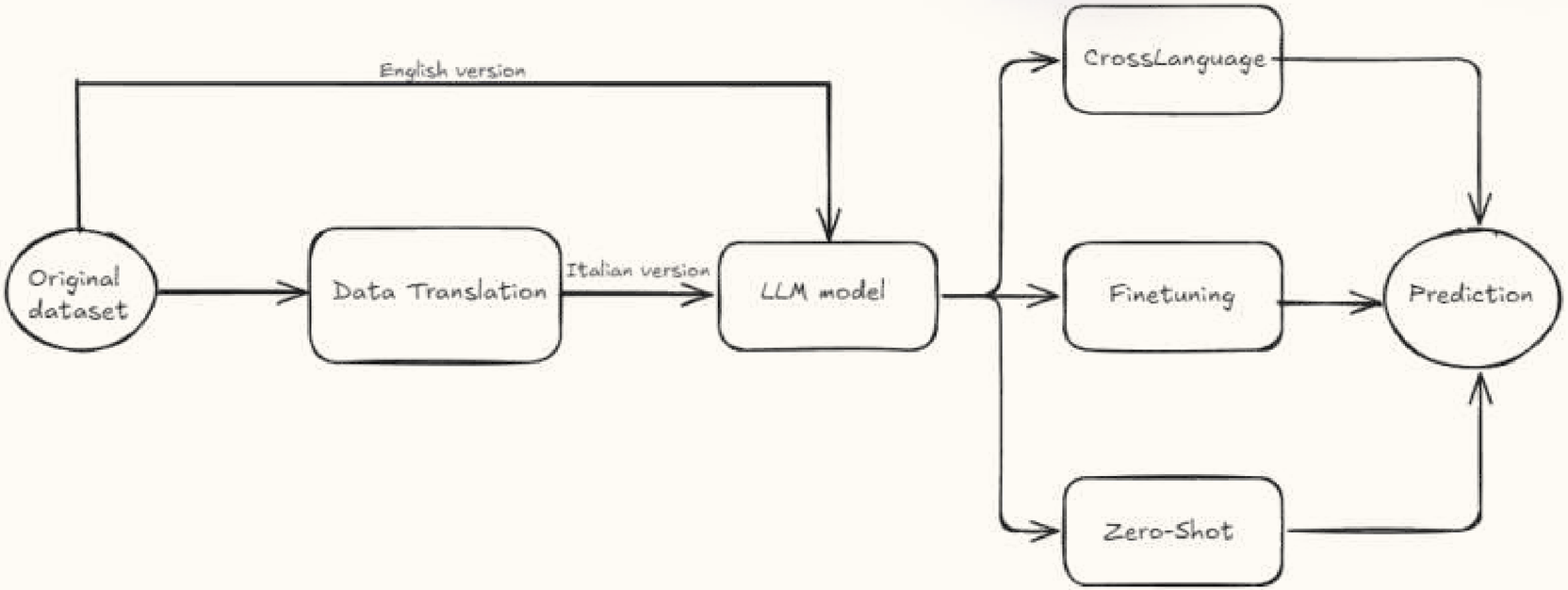
- **Dreaddit**
  - Designed to identifying **stress** in social media texts, with a focus on Reddit.
  - Binary classification

- **DepSeverity**
  - Created to assist in early identification and classification of **depression severity** levels through user-generated content from Reddit.
  - Binary and multicass classification (Minimum, Moderate, Mild and Severe)

- **SDCNL**
  - Contains posts from individuals expressing **suicidal intentions**

Macro-schema and Models

- **Bert**
  - Crosslanguage
  - Finetuning
  - Zero-Shot
- **RoBERTa**
  - Crosslanguage
  - Finetuning
- **Flan-T5**
  - Crosslanguage
  - Finetuning
  - Zero-Shot
- **Qwen 0.5B**
  - Crosslanguage
  - Finetuning
  - Zero-Shot
- **Qwen 1.5B**
  - Zero-Shot

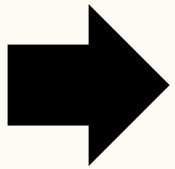


Data Translation

Helsinki-NLP/opus-mt-en-it

- Developed by the **Helsinki-NLP team**,
- Trained on a large dataset of parallel data, pairs of English and Italian sentences.
- **Encoder** that processes the English text
- **Decoder** that generates the translation in Italian.

And then I just want it to go away. I dont know what to do about this, I feel like the world doesn't want me to talk about it. It's to uncomfortable for people... But I hate these feelings and they are just the tip of the iceburg that is ptsd. I felt like I needed to say something somwhere though.



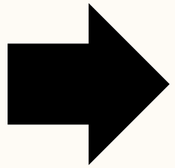
Non so cosa fare a riguardo, mi sento come se il mondo non volesse che ne parlassi. E' a disagio per le persone... ma odio questi sentimenti e sono solo la punta dell'iceburg che `e ptsd. Mi sentivo come se avessi bisogno di dire qualcosa somwhere

Data Translation

GoogleTranslator

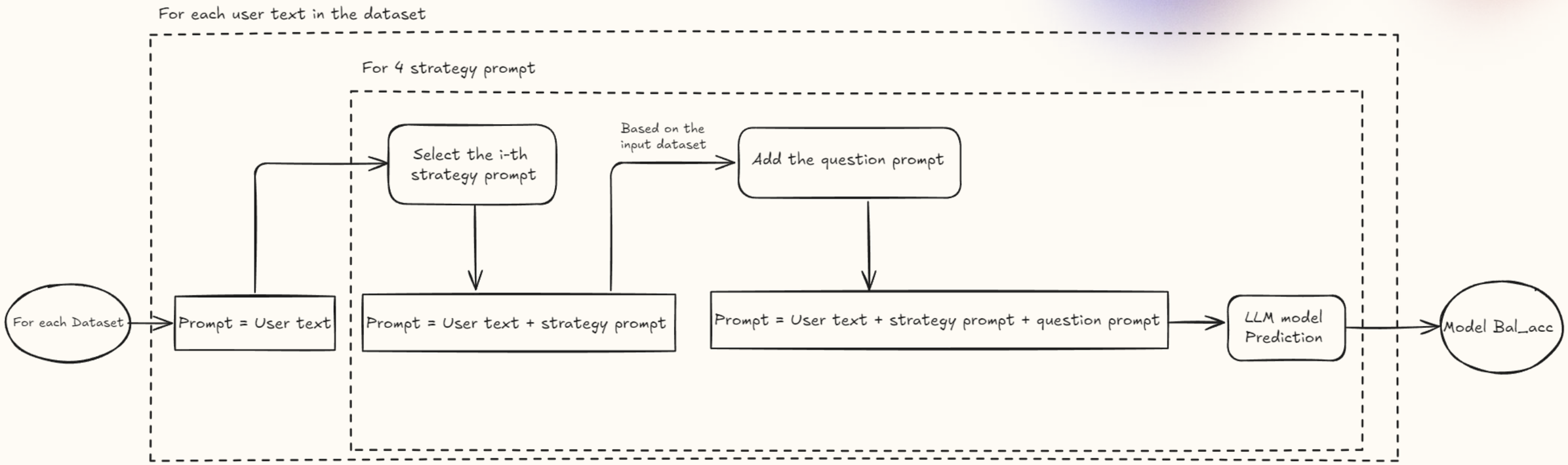
- Automatic translation service developed by Google.
- **Neural network models**, to deliver **real-time translations**
- **Deep-translator** library, to calls Google Translator APIs.

And then I just want it to go away. I dont know what to do about this, I feel like the world doesn't want me to talk about it. It's to uncomfortable for people... But I hate these feelings and they are just the tip of the iceburg that is ptsd. I felt like I needed to say something somwhere though.



E poi voglio solo che se ne vada. Non so cosa fare al riguardo, sento che il mondo non vuole che ne parli. E' troppo scomodo per le persone... Ma odio questi sentimenti e sono solo la punta dell'iceberg del disturbo da stress post-traumatico. Però sentivo il bisogno di dire qualcosa da qualche parte.

Schema for Zero-Shot Prediction



Strategy Prompts

Strategy	Prompt
Basic	
Context Enhancement	Considera questo post presente sui social media per rispondere alla seguente domanda.
Mental Health Enhancement	Come se fossi uno psicologo, leggi il post scritto su un social media e rispondi alla domanda.
Context & Mental Health Enhancement	Questa persona ha scritto questo paragrafo su un social media. Come se fossi psicologo, considera la condizione di benessere mentale espressa in questo post, leggi il post scritto su un social media e rispondi alla domanda.

Question Prompts

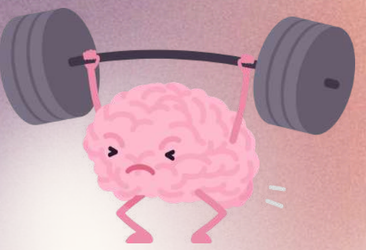
Dataset	Prompt
Dreddit	Lo scrittore di questo post risulta essere stressato?Rispondi solo con Sì o No.
SDCNL	Lo scrittore di questo post vuole suicidarsi? Rispondi solo con Sì o No
DepSeverity	Lo scrittore di questo post risulta essere depresso? Rispondi solo con Sì o No.
DepSeverity (Multiclass)	A che livello lo scrittore di questo post risulta essere depresso? Rispondi con un livello da 0 a 3 , dove 0 `e il minimo.

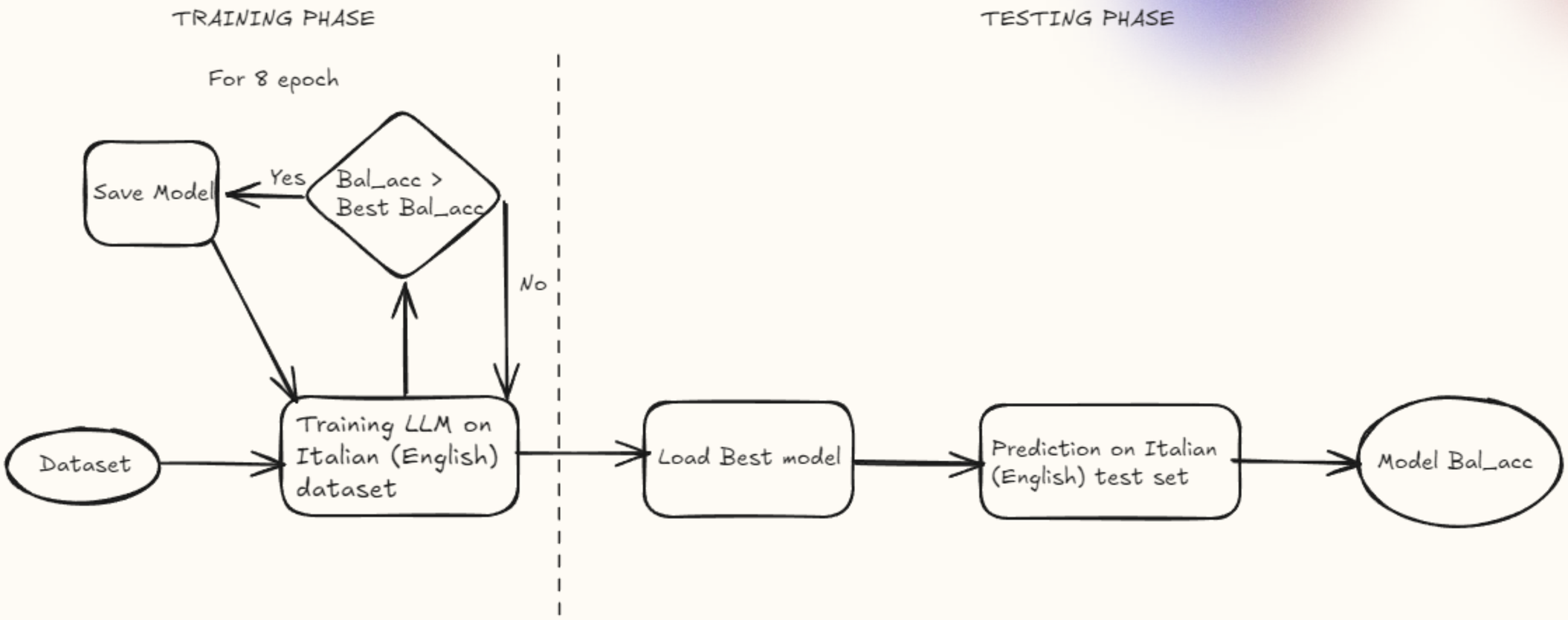
Results Zero-Shot Prediction

	Qwen2-0.5B	Qwen2 1.5B	BERT	Flan-T5
Dreaddit	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.591	0.654	0.504	**0.692** (0.598)
Italian (Google)	0.525	**0.695**	0.505	0.662
Dep_severity	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.524	**0.527**	0.505	0.510 (0.501)
Italian (Google)	0.502	**0.520**	0.508	**0.520**
Dep_severity_task 2	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.594	**0.729**	0.563	0.715 (0.682)
Italian (Google)	0.533	**0.686**	0.524	0.636
SDCNL	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.541	**0.663**	0.535	0.656 (0.589)
Italian (Google)	0.555	**0.591**	0.514	0.523

Strategy	Percentage
Basic	18,7%
Context Enhancement	9,3%
Mental Health Enhancement	18,7%
Context & Mental Health Enhancement	53,1%

Schema for Fine-tuning prediction

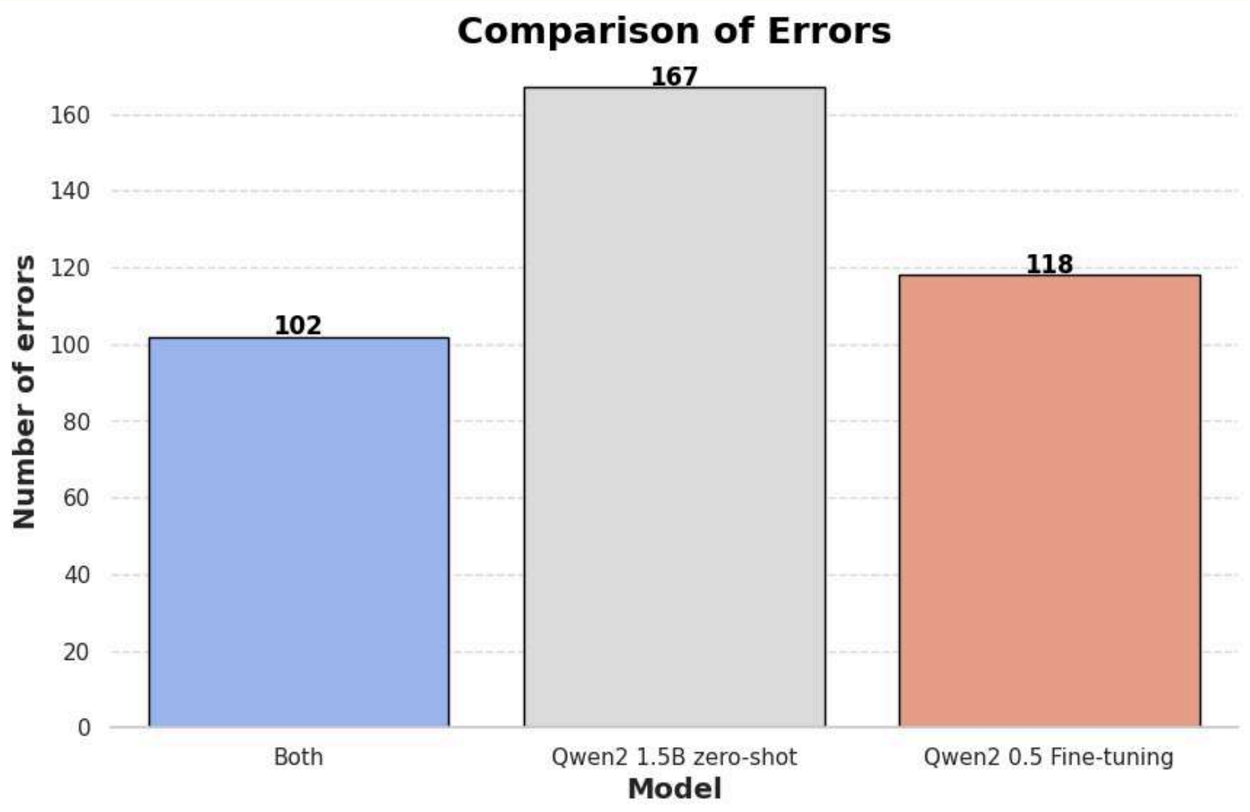




Results Fine-tuning Prediction

	RoBERTa	BERT	Flan-T5	Qwen2 0.5B
Dreaddit	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	**0.817**	0.800	0.796	0.677
Italian (Google)	0.717	**0.790**	0.726	0.697
Italian (Helsinki)	0.750	**0.759**	0.733	0.628
Dep_severity	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.696	0.716	**0.723**	0.593
Italian (Google)	0.656	**0.693**	0.657	0.551
Italian (Helsinki)	**0.668**	0.667	0.660	0.555
Dep_severity_task 2	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.778	**0.794**	0.791	0.726
Italian (Google)	0.7318	**0.785**	0.730	0.657
Italian (Helsinki)	0.7242	**0.780**	0.737	0.638
SDCNL	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English (Original)	0.663	**0.757**	0.721	0.658
Italian (Google)	0.660	**0.681**	0.655	0.626
Italian (Helsinki)	0.642	**0.689**	0.645	0.577

Fine-tuning vs Zero-Shot



- **Qwen2 0.5B fine-tuned**: 220 errors, better performance despite fewer parameters.
- **Qwen2 1.5B zero-shot**: 269 errors without fine-tuning.

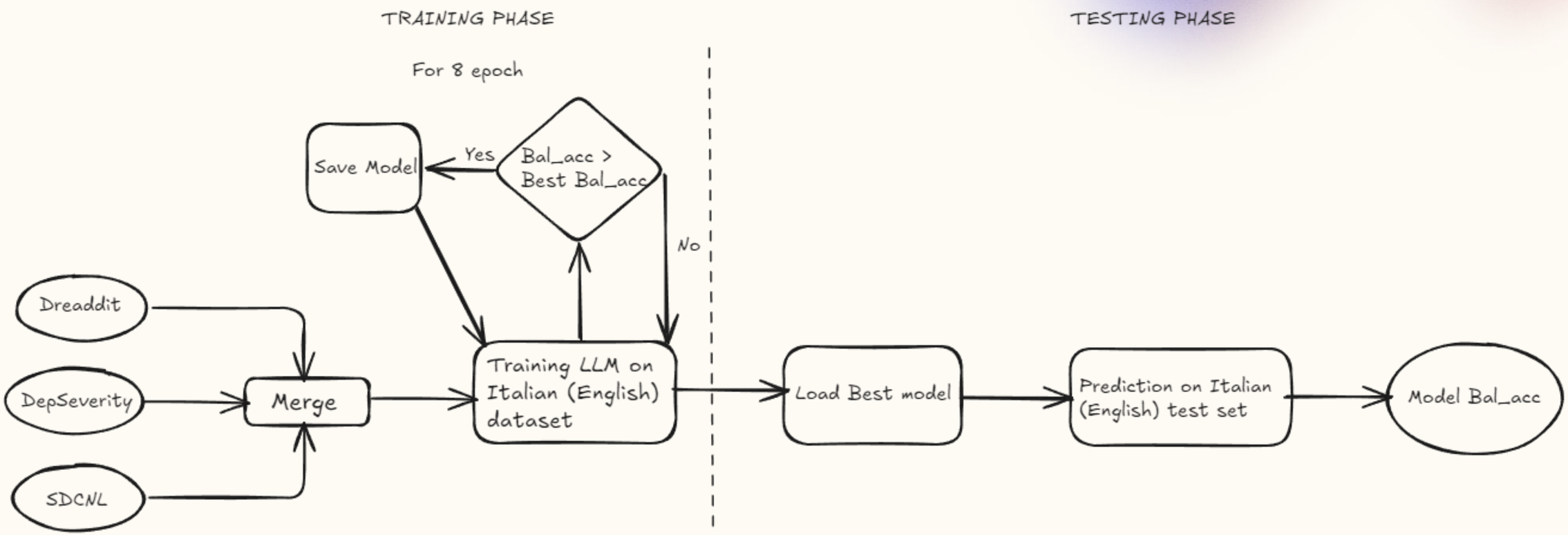
Fine-tuning has significantly improved accuracy, even for a smaller model. A finetuned model can outperform a larger zero-shot model, reducing computational costs.

Results CrossLanguage Prediction

	RoBERTa	BERT	Flan-T5	Qwen0.5
Dreaddit	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English-Italian(Google)	**0.759**	0.689	0.528	0.504
Italian(Google)-English	0.773	0.735	**0.781**	0.502
Italian(Helsinki)-English	0.744	0.715	**0.770**	0.519
Dep_severity	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English-Italian(Google)	0.610	0.565	**0.701**	0.514
Italian(Google)-English	0.500	**0.642**	0.520	0.517
Italian(Helsinki)-English	0.500	**0.634**	0.600	0.511
Dep_severity_task 2	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English-Italian(Google)	**0.752**	0.704	0.553	0.535
Italian(Google)-English	**0.774**	0.682	0.714	0.560
Italian(Helsinki)-English	**0.767**	0.717	0.713	0.615
SDCNL	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**	**Bal_Acc**
English-Italian(Google)	**0.686**	0.567	0.657	0.508
Italian(Google)-English	0.658	0.664	**0.680**	0.506
Italian(Helsinki)-English	0.589	**0.654**	0.590	0.501

Zero-Shot and **CrossLanguage** prediction are not comparable to the performance achieved after **Fine-tuning**

Schema for Italian-Mental-LLM



Results for Italian-Mental-LLM

Model	Bal_Acc
Italian-Mental-Flan-T5	0.728
Italian-Mental-BERT	0.757
Italian-Mental-RoBERTa	0.745
Italian-Mental-Qwen2 0.5B	0.658

Model	Bal_Acc
English-Mental-Flan-T5	0.771
English-Mental-BERT	0.767
English-Mental-RoBERTa	0.767
English-Mental-Qwen2 0.5B	0.725

Conclusion

The study introduces three Italian Mental Health datasets and four models capable of predicting **stress**, **anxiety**, **depression**, and **suicidal tendencies** in **Italian**, marking a significant advance in mental health research in Italy.

Introduce new Italian mental health datasets is essential to improve **model generalization** by incorporating more annotated data and collaborating with **mental health professionals**.

Upgrading **computational resources** would enable the use of larger models, allowing for more **accurate predictions**.

Thank you for your attention!