universidad de león

ALBA

Deep Learning–Based Hospital Admission Prediction from Spanish Psychiatric Electronic Health Records

Arturo Crespo-Álvaro[1], Sergio Rubio-Martín[1], María Teresa García-Ordás[1], Antonio Serrano-García[2], Clara Margarita Franch-Pato[2], Alicia Merayo-Corcoba[1] and José Alberto Benítez-Andrades[1]

[1] *ALBA Research Group, Department of Electric, Systems and Automatics Engineering, Universidad de León, Spain.*

[2] *Servicio de Radiología, Complejo Asistencial Universitario de León (CAULE), Spain.*

(June, 2026)

acrea@unileon.es

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

AGENCIA ESTATAL DE INVESTIGACIÓN

PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in HER (SUICIDETECT)”
MICIU/AEI/10.13039/501100011033

universidad de león

ALBA

Table of Contents

1 Introduction

2 State of the Art

3 Dataset Description

4 Methodological pipeline

5 Models & Dataset Split

6 Solution Overview & Optimization

7 Results

8 Conclusions & Future Work

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES AGENCIA ESTATAL DE INVESTIGACIÓN PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

universidad de león

ALBA

About us:

León, Spain

In the northwest of Spain, a region with a strong tradition in innovation, education and quality of life.

- A strategic location well connected to major Spanish and European cities.
- **A dynamic university city** with a rich cultural heritage and a vibrant community.
- A hub for talent, research and technological transfer.

UNIVERSITY OF LEÓN

A public university with a strong commitment to teaching, research and innovation.

- High-quality education and research
- Excellence in Engineering and Data Science
- Interdisciplinary collaboration
- International outlook and impact

ALBA LAB

We develop and apply **Artificial Intelligence** to solve real healthcare challenges and improve people's lives.

CLINICAL NLP
Extracting knowledge from clinical texts and EHRs.

PREDICTIVE ANALYTICS
Building models to predict outcomes and support decisions.

KNOWLEDGE GRAPHS
Structuring and linking biomedical knowledge.

CLINICAL DECISION SUPPORT
Developing intelligent tools to support healthcare professionals.

OUR MISSION
Transforming clinical data into actionable medical knowledge through AI, to enable better decisions and better care.

CLINICAL DATA → AI MODELS → DECISION SUPPORT → HEALTHCARE IMPACT

- **100+** PUBLICATIONS
- **50+** CONFERENCES
- NATIONAL & INTERNATIONAL COLLABORATIONS
- COMPETITIVE RESEARCH PROJECTS
- REAL-WORLD IMPACT

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES · AGENCIA ESTATAL DE INVESTIGACIÓN

PID2023-146168OA-I00 "NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)" MICIU/AEI/10.13039/501100011033

universidad de león

ALBA

About us: ALBA members

Mª Teresa García Ordás

José Alberto Benítez Andrades

Isaías García Rodríguez

Alicia Merayo Corcoba

Sergio Rubio Martín

Arturo Crespo Álvaro

LinkedIn

@ALBALAB_EU

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
AGENCIA ESTATAL DE INVESTIGACIÓN
PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

Introduction

1 in 7 people

Living with a mental disorder worldwide

≈ 1.1 billion people (WHO 2021)

High clinical risk

Severe psychiatric disorders are associated with:

- Increased suicide risk
- Physical comorbidities
- 10–20 years reduced life expectancy

Limited access to care

Long waiting times and shortage of specialists can delay treatment and increase hospitalization risk.

Documentation Complexity

Spanish psychiatric notes present additional challenges (**figurative language**, **abbreviations**, **inconsistency**).

Narrative Language

Extracting structured data from Spanish psychiatric notes presents **unique linguistic and domain-specific difficulties**.

Main Goal

Early and objective identification of **high-risk patients** can **optimize healthcare resources** and **improve the decision-making process**.

PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

State of the art

How can we automatically identify which patients need to be hospitalized based on their clinical notes?

- NLP in medicine has advanced in **automatic document classification**, risk **prediction**, and analysis of Electronic Health Records (EHRs).
- **Deep Learning** models and **Transformer-based architectures** (BERT, EHR-BERT, BiLSTM) outperform traditional methods in clinical prediction and medical text analysis.
- In psychiatry, AI has been applied to **text classification**, clinical entity **extraction and normalization**, and prediction of mental health crises using multimodal approaches.
- Mental health–specific models, such as **Guardian-BERT**, can detect subtle **linguistic signals** associated with clinical risk (e.g., suicidal ideation).
- Research on **Spanish psychiatric texts** is still limited, and no studies were found on **hospitalization prediction** using unstructured clinical notes with deep learning models.
- This work addresses this gap by **evaluating advanced NLP models** to automatically classify psychiatric clinical notes based on the need for patient hospitalization.

PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

Dataset information

DATA EXAMPLE

MC 1 | APC 2 | APF 3 | DX 4 | TTF 5 | TTA 6

Mujer de 25 años diagnosticada de esquizofrenia simple con varios ingresos en Unidad de Cuidados Especiales (último en 2013). Seguimiento en Centro de Salud Médico de José Aguado en tratamiento con leponex 100(0-0-1/2), abilify 10 miligramos (1-0-0) y escitalopran 1 (0-1/2-0). Acude a urgencias refiriendo oir voces de 5 deías de evolución, en forma de insultos y amenazas. ESTADO PSICOLÓGICO DEL PACIENTE: consciente, orientada. escasamente colaboradora. Inhibición psicomotriz. Mal contacto visual. Discurso pobre con aumento numero latencia de respuesta. Alucinaciones auditivas. Ideación delirante autorreferencial. No ansiedad-No ideas de muerte. No autoagresividad/heteroagresividad. IMPRESIÓN DIAGNÓSTICA: ESQUIZOFRENIA SIMPLE PLANIFICACIÓN: Abilify 10 miligramos: 1 comprimido por la mañana y otro medio en la comida Heipram 10 miligramos: 1 comprimido a la hora de la comida. Resto tratamiento igual Consulta preferente en Centro de Salud Médico. En caso de no mejoría volver a urgencias.

Data source:

Complejo Asistencial Universitario de León (CAULE).

Original state

500 clinical notes from psychiatric emergencies in **free text without structure** and completely anonymized data.

Preprocessing stage

Cleaning, spelling correction and **manual NER labeling** to ensure the quality of the set.

Final result

409 valid records after the debugging process, ready for model training.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

AGENCIA ESTATAL DE INVESTIGACIÓN

PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

NER Entities Identified in the Data

Clinical notes were annotated with six categories of key clinical entities to structure patient information.

Chief Complaint (MC):

Main reason for the psychiatric consultation.

Clinical History (APC):

Relevant past medical or psychiatric conditions (diagnoses, surgeries...).

Pharmacological History (APF):

History of previous treatments with medications and other substances.

Diagnosis (DX) :

Psychiatric disorder(s) identified by the clinician according to guidelines.

Pharmacological Treatment (TTF):

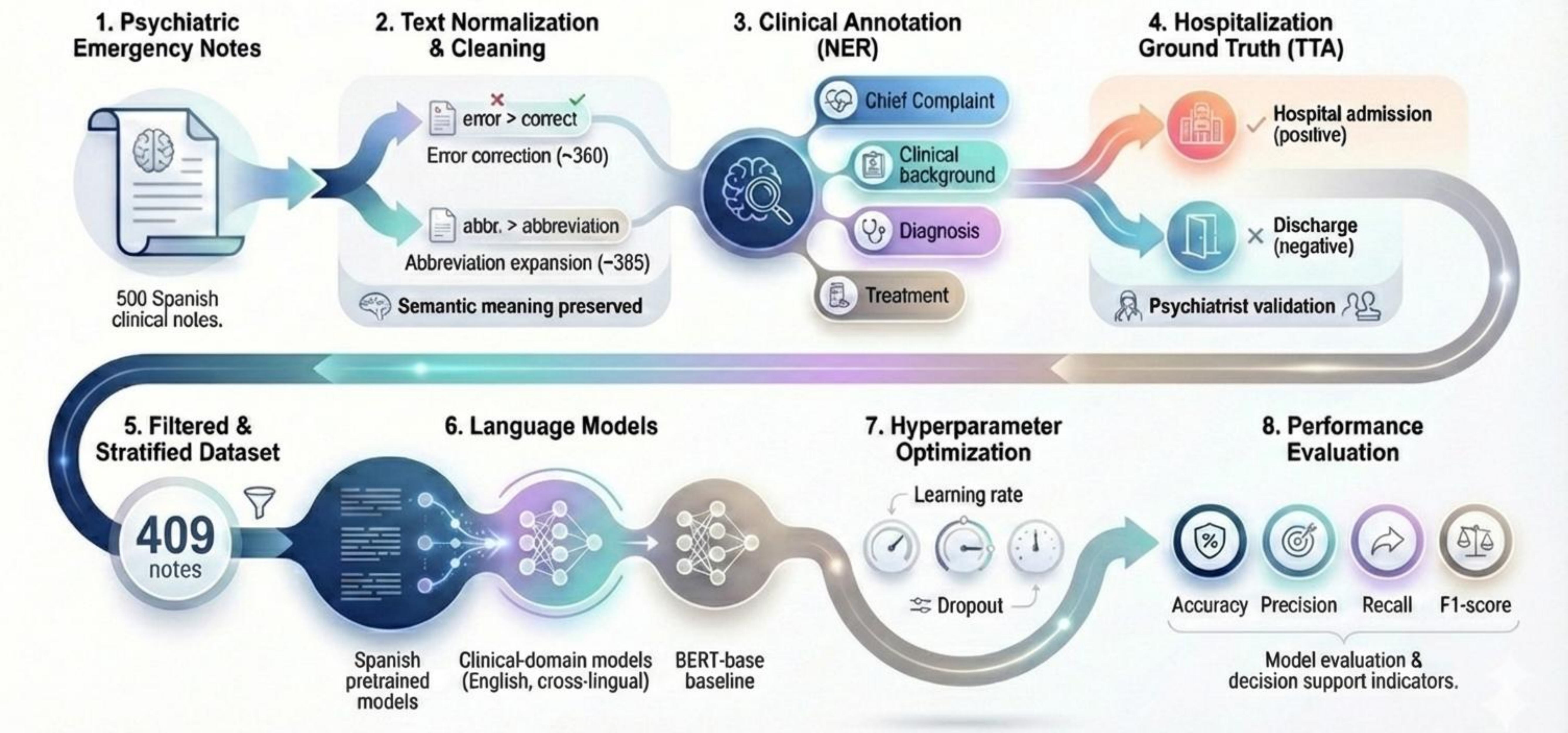
Medications currently prescribed for psychiatric treatment.

Action-Based Treatment (TTA):

Non-pharmacological interventions, such as hospitalization or discharge.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
AGENCIA ESTATAL DE INVESTIGACIÓN
PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

Methodological pipeline



PID2023-146168OA-I00 "NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)" MICIU/AEI/10.13039/501100011033

Models & Dataset split

BERT Base

BERT base model pretrained with a general English corpora. Available **cased** and **uncased** versions

BERT Beto

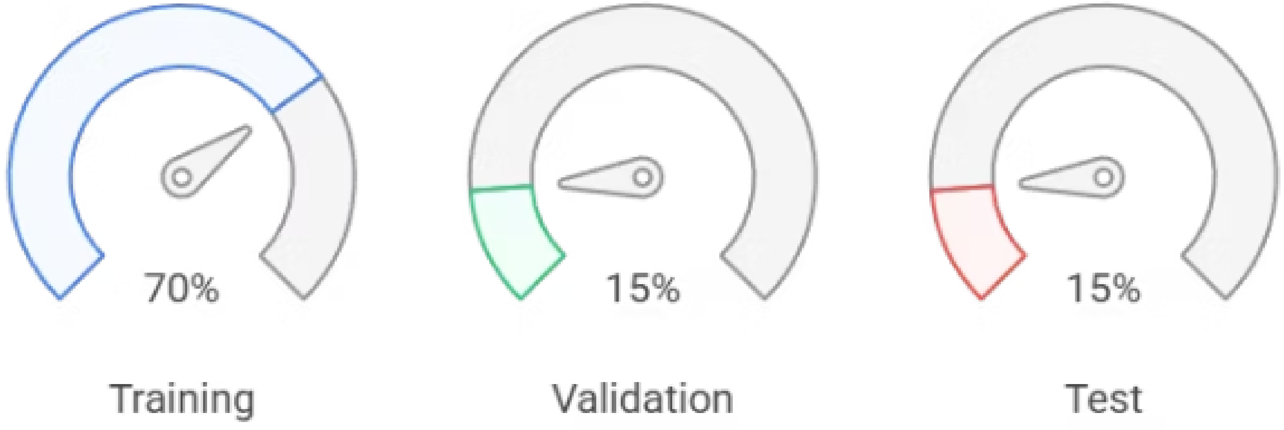
BERT base model pretrained with a general Spanish corpora. **cased** and **uncased** variants

Clinical BERT

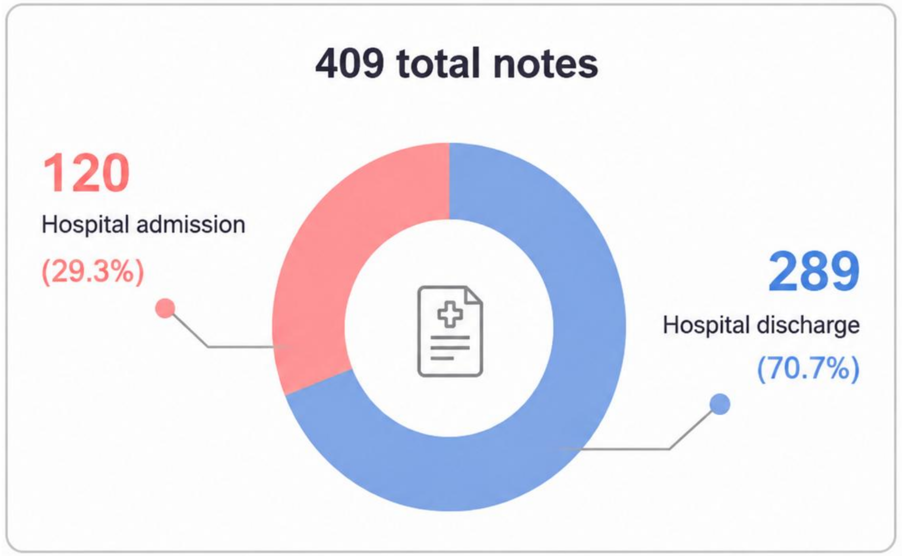
Domain specific BERT pretrained with EHRs and without **cased** or **uncased** versions.

Dataset distribution

Stratified split to maintain the proportion of samples from each class among the fine-tuning process.



Data imbalance



Hospital admission	120 notes	29.3%
Hospital discharge	289 notes	70.7%

PID2023-146168OA-I00 "NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)" MICIU/AEI/10.13039/501100011033

Solution Overview & Optimization

Main Approach

Classification using the full text of the clinical note as written, without additional structural processing

Baseline

Using the models with their default configuration, such as learning rate, train epochs for the fine-tunning.

Grid Search

Testing different configurations for each model in order to obtain the best results in the fine-tuning process.

Hyperparameter Search Space

Values explored during results optimization.

Hyperparameter	Values Explored
Batch size (per device)	{8, 16}
Learning rate	$\{2e^{-5}, 1e^{-5}, 3e^{-5}\}$
Number of epochs	{3, 4}
Weight decay	{0.0, 0.01}

PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)” MICIU/AEI/10.13039/501100011033

universidad de león

ALBA

Results overview

Model performance comparison

Comparison of the models using two approaches: default configuration and optimized configuration (With Hyperparameters).

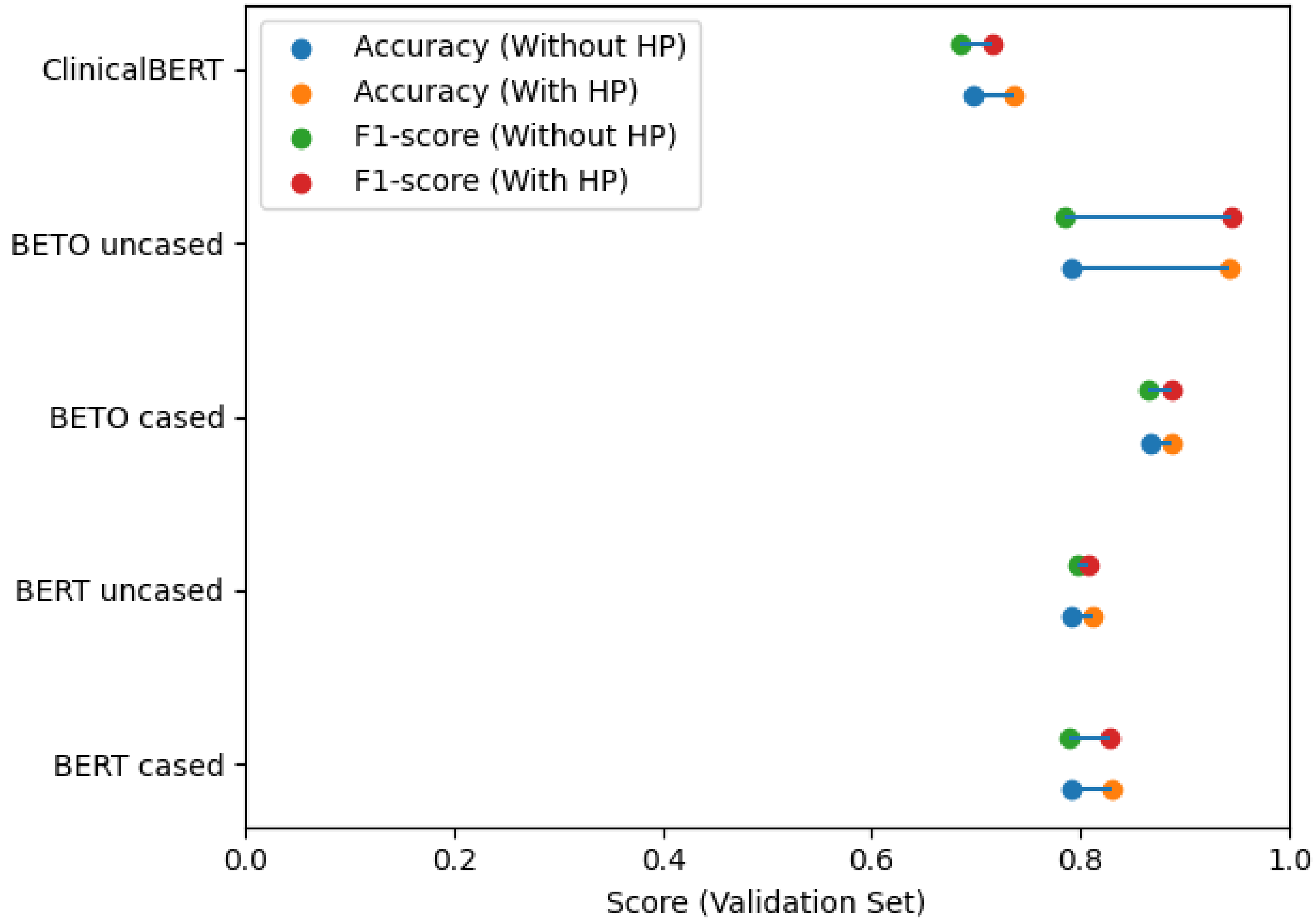
Metrics reported on validation and test sets:

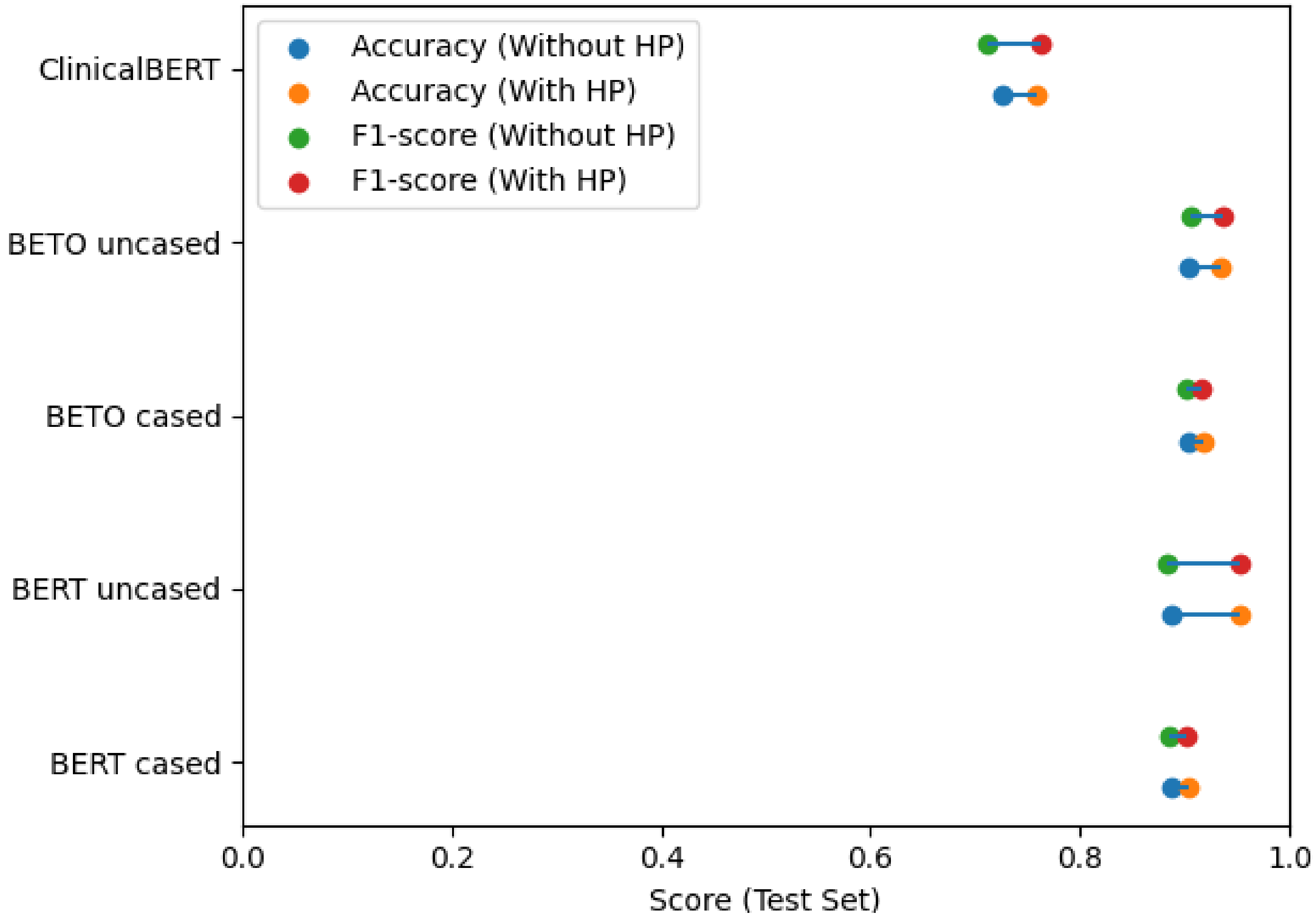
- Accuracy: proportion of correct predictions.
- F1-score: harmonic mean between precision and recall.

Model	Method	Validation		Test	
		Accuracy (higher is better)	F1-score (higher is better)	Accuracy (higher is better)	F1-score (higher is better)
Clinical BERT	Default configuration	69.8%	68.4%	72.6%	71.1%
	With Hyperparameters	**73.6%**	**71.5%**	**75.8%**	**76.3%**
BETO (uncased)	Default configuration	79.2%	78.6%	90.3%	90.5%
	With Hyperparameters	**94.3%**	**94.4%**	**93.5%**	**93.6%**
BETO (cased)	Default configuration	86.8%	86.4%	90.3%	90.1%
	With Hyperparameters	**88.7%**	**88.7%**	**95.2%**	**95.2%**
BERT (uncased)	Default configuration	79.2%	79.7%	88.7%	88.4%
	With Hyperparameters	**83.0%**	**82.5%**	**93.5%**	**93.5%**
BERT (cased)	Default configuration	79.2%	79.0%	88.7%	88.6%
	With Hyperparameters	**83.0%**	**82.9%**	**90.3%**	**90.1%**

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
AGENCIA ESTATAL DE INVESTIGACIÓN
PID2023-146168OA-I00 "NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)" MICIU/AEI/10.13039/501100011033

Results





- The graph on the left compares the performance of the validation set, while the graph on the right compares the performance of the test set.

PID2023-146168OA-I00 "NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)" MICIU/AEI/10.13039/501100011033

Conclusions & Future Work

Key findings

1. Hyperparameter optimization consistently improves performance

Particularly in BETO uncased, with large gains in validation (Accuracy: 0.792 → 0.943; F1-score: 0.786 → 0.944).

2. BETO cased with optimization achieves the best test performance

It reaches the highest results on the independent test set (Accuracy: 0.952; F1-score: 0.952), showing strong generalization.

3. Language alignment is crucial

Spanish-pretrained models (BETO, BERT) outperform ClinicalBERT, indicating that linguistic alignment has more impact than domain pretraining alone.

4. Transformer models are effective

After tuning, BETO uncased, BERT uncased and BETO cased surpass 0.93 in both Accuracy and F1-score on the test set.

Limitations

- **Relatively small dataset:** Although carefully curated, the dataset size (409 EHRs) may limit model capacity.
- **Single-center data:** All notes come from a single institution, which may restrict generalizability.
- **Free-text and binary classification only:** The study does not include structured variables or multi-class scenarios.

Future work

1. **Expand to multi-center datasets**
   Evaluate models across different institutions to assess cross-institutional robustness and generalizability.
2. **Explore hybrid architectures**
   Combine structured clinical variables with unstructured text to leverage complementary information.
3. **Evaluate ensemble strategies**
   Investigate ensemble and stacking approaches to enhance stability and predictive performance.
4. **Extend to more complex scenarios**
   Study multi-class predictions (e.g., length of stay, discharge destination) and early risk stratification models.

PID2023-146168OA-I00 "NLP-Driven Insight Engine for Suicide Attempt Detection in EHR (SUICIDETECT)" MICIU/AEI/10.13039/501100011033

Deep Learning–Based Hospital Admission Prediction from Spanish Psychiatric Electronic Health Records

Arturo Crespo-Álvaro[1], Sergio Rubio-Martín[1], María Teresa García-Ordás[1], Antonio Serrano-García[2], Clara Margarita Franch-Pato[2], Alicia Merayo-Corcoba[1] and José Alberto Benítez-Andrades[1]

[1] *ALBA Research Group, Department of Electric, Systems and Automatics Engineering, Universidad de León, Spain.*

[2] *Servicio de Radiología, Complejo Asistencial Universitario de León (CAULE), Spain.*

(June, 2026)

acrea@unileon.es

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

AGENCIA ESTATAL DE INVESTIGACIÓN

PID2023-146168OA-I00 “NLP-Driven Insight Engine for Suicide Attempt Detection in HER (SUICIDETECT)”
MICIU/AEI/10.13039/501100011033