Predicting Suicide Attempts with a Long-Short Term Memory Using Environmental Data

Pol Capdevila, Dolores Rexachs and Jordi Cahué

July 15, 2024

Predicting suicide attempts with a long-short term memory using environmental data

Pol Capdevila[1,2][0009−0006−9643−6539], Dolores Rexachs[1][0000−0001−5500−850X], and Jordi Cahué[2]

[1] Universitat Autònoma de Barcelona
[2] Kyndryl Spain

Abstract. Hospitals are presently in need of predictive tools to anticipate emergent situations, particularly those pertaining to mental health crises such as suicide attempts. Despite prior research indicating potential influencing factors, the development of effective solutions remains a formidable challenge. However, recent technological advancements, particularly in artificial intelligence (AI), offer promising avenues for addressing this challenge. Building on these advancements, the present study develops and trains a predictive model utilizing a Long Short-Term Memory (LSTM) neural network. The model is trained using data on suicide attempt admissions and environmental variables, as an influence factor, from a hospital in Catalonia. Results demonstrate the potential of AI to provide valuable insights to hospitals, aiding in the management and optimisation of healthcare resources to effectively address mental health emergencies.

Keywords: S · LSTM · Mental health · Suicide attempts

1 Introduction

Human beings have instinctively associated environmental conditions with health outcomes, exemplified by seeking shelter during inclement weather or cooler temperatures to evade cold-related illnesses—a rudimentary form of survival instinct. This innate proclivity for survival has served as a driving force for disease prevention and learning. For instance, residing in areas characterised by elevated air pollution levels has been linked to heightened susceptibility to lung cancer [13].

In mental health, investigations have revealed associations between environmental phenomena and psychiatric outcomes. Studies have demonstrated a correlation between heat waves and heightened psychiatric admissions [2], indicating the impact of temperature extremes on mental health crises. Moreover, empirical evidence suggests that temperature fluctuations exert an influence on suicide rates among human populations [11], underscoring the multifaceted interplay between climatic conditions and mental well-being. Despite these strides, the integration of Artificial Intelligence (AI) into mental health research remains relatively nascent, with advancements in this domain lagging behind those observed in fields such as finance or imaging for cancer detection, indicating a need for further development and exploration in leveraging AI for mental health applications.

Hospital admissions due to suicide attempts represent a critical facet of mental health care, reflecting the consequences of untreated psychiatric conditions and societal stressors. Understanding the patterns and determinants of hospital admissions for suicide attempts is crucial for informing targeted interventions and resource allocation strategies aimed at reducing suicide rates. This research utilises a time-series analysis framework to uncover temporal dynamics and trends in behaviours.

Thus, adopting a time-series analysis perspective not only enhances comprehension of the temporal dynamics surrounding suicide attempts but also holds promise in guiding evidence-based interventions, aligning with the primary objective of this article to assist hospitals in resource management and optimisation.

The primary hypothesis behind this segmentation is that if the model accurately predicts high-demand days, the hospital can prepare and allocate resources accordingly. Conversely, if the model confirms the absence of high-demand days, resources can also be efficiently organized.

2 Methods and materials

Integrating environmental factors such as weather conditions and contamination levels with hospital admission data presents a multifaceted approach towards understanding the interplay between human health and the environment. By examining how variations in weather patterns intersect with hospital admissions, we can elucidate potential correlations and causal relationships between environmental exposures and health outcomes. This interdisciplinary approach not only elucidates the acute effects of environmental exposures on hospitalisation rates but also holds promise in informing proactive measures for mitigating adverse health effects and optimising healthcare resource allocation using AI.

2.1 Data

Although we had data starting in 2018, it was decided to only use post COVID-19 data in order to avoid any effect that the lockdown had in 2020. So, the data used in this study comprehend almost 2 years of records, from January 1st 2021 to August 31th 2022.

Admissions of suicide attempts The dataset employed in this study encompasses hospital admissions from a facility situated in Catalonia, catering to 25 adjacent municipalities, detailing daily admission frequencies. Notably, demographic attributes such as age and gender, as well as the patient's medical history, are excluded from consideration. Instead, the analysis focuses exclusively on identifying daily suicide attempts, while implementing measures to avoid gender bias.

Environmental factors Environmental factors are intertwined with human health, and numerous studies [10, 1] have delved into this relationship. In this research, we harnessed data sourced from monitoring stations dedicated to assessing pollution levels and meteorological conditions. These stations are under the jurisdiction of the Generalitat de Catalunya or Meteocat, the Catalan meteorological service, both of which offer readily accessible open data resources.

Weather data The variables accessible at the data collection stations encompassed temperature, humidity, atmospheric pressure, rainfall, solar irradiance, and wind velocity. In most studies, the increment of suicide attempts happen with extreme values of those variables, as suggested in [10] being the temperature the variable with most studies.

Pollution data While pollution data has been extensively explored in the context of its association with respiratory diseases [12, 7, 14], the literature concerning its impact on mental health remains comparatively limited. Nonetheless, there exists a body of research that studies the correlation between air pollution and self-harm, as evidenced by studies such as that presented in [1].

The variables employed comprised those accessible from public monitoring stations, encompassing PM2.5 (particulate matter with a diameter of 2.5 microns or less), PM10 (particulate matter with a diameter of 10 microns or less), as well as pollutants such as Plumb, Nickel, Cadmium, Arsenic, Carbon Monoxide, Nitrogen Monoxide, Sulfur Dioxide, and Ozone. The investigation into the influence of these variables on the predictive performance of the model was deferred to subsequent analyses.

Other variables In order to map the temporality in the model, we need to add variables that provide that information such as a binary variable if it is a vacation day or an integer variable having the day and the month. It also has been added variables that map public holidays and school holidays as shown in [4] where a possible correlation between them and suicide attempts was found.

In mental health, various determinants, including economic factors, can contribute to suicide risk as shown in [9, 8]. The study presented in this report does not incorporate individual-level data, but such information could be supplemented with indicators like municipal indebtedness or median income levels. Future research endeavours could undertake a more comprehensive investigation, integrating these socio-economic metrics to enhance our understanding of the complex interplay between economic circumstances and suicidal behaviour.

2.2 Data preparation

The initial challenge tackled in this research pertained to imperfections inherent in data collection systems, evidenced by a 4% incidence of missing values in the collected dataset. Notably, the predominant nature of these missing entries was characterised as punctual, occurring between two valid data points. Leveraging this observation, corrective measures were implemented via linear extrapolation to rectify the dataset, thereby mitigating the impact of missing values on subsequent analyses. An additional hypothesis posited was the universality of the envisioned system for hospital-wide implementation, implying its applicability across all towns simultaneously. To facilitate data preparation, a decision was made to compute the arithmetic mean of variable values, given the hospital's reference area of 30 kilometres. This approach, while potentially compromising predictive accuracy at the tails of variable distributions affords the development of a robust model capable of encompassing the hospital's reference area in a cohesive manner.

To create the training, validation, and test sets, the data was partitioned into three segments. The initial 70% of the data, chronologically ordered, was allocated for training the neural network, while the subsequent 10% was used for validation and the final 20% for testing. This partitioning methodology ensures the robustness and generalisability of the results to other time periods.

3 Experiments

In order to predict daily admissions, a Neural Network has been trained, specifically a Long-Short Term Memory (LSTM) introduced in [5] which represents a sophisticated variant of Recurrent

Neural Networks (RNNs) designed to overcome the limitations of traditional RNNs [6]. Unlike standard RNNs, LSTM networks incorporate specialised memory cells equipped with gating mechanisms, enabling them to selectively retain and propagate information over extended time horizons. This architecture empowers LSTM to effectively learn temporal dependencies while mitigating the vanishing gradient problem commonly encountered in traditional RNNs. Consequently, LSTM networks have found widespread applications in various fields, including natural language processing, speech recognition, time series forecasting, and medical diagnosis, owing to their ability to model complex sequential patterns and exhibit robust performance in tasks requiring long-range dependencies.

3.1 Time forecasting and LSTM architecture

To select an appropriate hyperparameter for forecasting future demand, various tests were conducted to assess its potential impact. The results indicated minimal variation in the values, leading to the decision to set the hyperparameter to 7 days, is it to say, a week for subsequent experiments. The best results were achieved with a two-layer LSTM, with each layer containing 20 neurons. Increasing the number of layers or neurons led to overfitting and poor generalization on the test data.

3.2 Predicting the number of daily admissions

The first experiment done was to train a LSTM to predict the number of daily admissions. The best results shown a mean absolute error (MAE) of 14.01 with a 16.65% of mean absolute percentage error (MAPE). The results exhibit a strong alignment with previous research [3], evidenced by comparable MAPE values.

3.3 Adding admission ranks

A second set of experiments involved modifying the target variable. Rather than predicting the number of daily admissions, two admission categories were created to explore the impact of altering the prediction target. This change in the predicted variable aims to better align with the model's objective of optimising hospital resources. By segmenting admissions into different ranks based on resource requirements, the model can anticipate and allocate resources accordingly.

The experiments made focused on binary classification, where one category represented high-demand and the other regular-demand. The subsequent experiment inverted this, with one category representing low-demand and the other regular-high demand. In both cases, the experiments yielded an AUC (Area Under the Curve) of 63%, as shown in Figure 1, validating the hypothesis that AI can provide value in this field. By setting a threshold with the training data, the model achieved a 100% accuracy in identifying the corresponding important rank in the testing set. This indicates that when the model forecasts a high-demand day, the likelihood of prediction failure is nearly negligible allowing a correct classification of approximately 11% of high-demand days as shown in Table 1 . Similar results were observed in the second experiment, where the model correctly classified nearly 9% of low-demand days as shown in Table 2. This error-free classification is crucial in this use case because the hospital must ensure adequate resources for each day; thus, the model must be flawless in predicting these low-demand days.

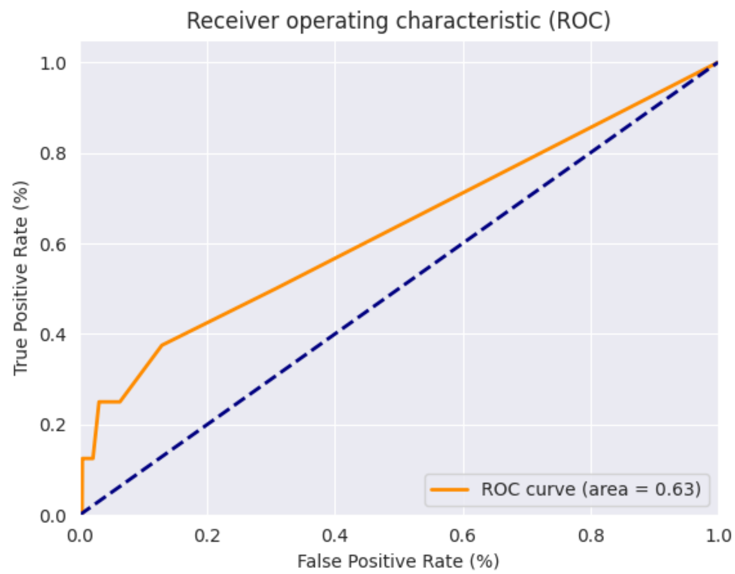

Fig. 1. ROC Curve of the first experiment

Table 1. Confussion Matrix High-Demand case

		Predicted Positive	Predicted Negative
Actual	Positive	3	25
	Negative	0	93

Table 2. Confussion Matrix Low-Demand case

		Predicted Positive	Predicted Negative
Actual	Positive	5	48
	Negative	0	68

4 Conclusions

While forecasting the flow of mental health emergencies is a complex task, this research has observed that AI can offer assistance. The models obtained, although not highly accurate, highlight this potential contribution. It's important to note that this study only evaluated one type of model and did not compare it with others; instead, it was compared with a similar study, yielding comparable results. On the other hand, analyzing the results of resource optimization based on admission ranks reveals that achieving error-free classification for approximately 10% of the days, could lead to a modest improvement in hospital resource allocation compared to the current practice, which lacks a systematic methodology, even though it is not possible to classify the remaining 90% of high-demand or low-demand days without error. These experiments should be further investigated with more data to conclusively validate the results obtained in order to create a robust model.

Acknowledgments I am grateful to the Generalitat of Catalonia for funding this study and to Kyndryl, and the Universitat Autònoma de Barcelona (UAB) for providing the opportunity to conduct this research.

Bibliography

[1] Aguglia, A., Giacomini, G., Montagna, E.: Meteorological variables and suicidal behavior: Air pollution and apparent temperature are associated with high-lethality suicide attempts and male gender. Front. Psychiatry 12:653390 (2021)

[2] Basagaña, X., Sartini, C., Barrera-Gómez, J., Dadvand, P., Cunillera, J., Ostro, B., Sunyer, J., Medina-Ramón, M.: Heat waves and cause-specific mortality at all ages. Epidemiology **22**(6), 765–772 (Nov 2011). https://doi.org/10.1097/ede.0b013e31823031c5, http://dx.doi.org/10.1097/EDE.0b013e31823031c5

[3] unified machine learning approach to time series forecasting applied to demand at emergency departments, A.: A unified machine learning approach to time series forecasting applied to demand at emergency departments. BMC Emerg Med 21 (2021)

[4] G. Jessen, B. F. Jensen, E.A.: Attempted suicide and major public holidays in europe: findings from the who/euro multicentre study on parasuicide. Acta Psychiatrica Scandinavica (2007)

[5] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (Nov 1997). https://doi.org/10.1162/neco.1997.9.8.1735, http://dx.doi.org/10.1162/neco.1997.9.8.1735

[6] Hopfield, J.: Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A. (1982)

[7] Kim, D., Chen, Z., Zhou, L.F., Huang, S.: Air pollutants and early origins of respiratory diseases. Chronic Diseases and Translational Medicine (2018)

[8] Kim, J., Kim, J., Choi, Y., Lee, T.H., Park, E.C.: Effect of socioeconomic status on the linkage between suicidal ideation and suicide attempts. Suicide and Life-Threating behaviour (2017)

[9] Michelle Funk, N.D., Knapp, M.: Mental health, poverty and development. Journal of Public Mental Health **11**(4), 166–185 (2012)

[10] Ngu, F., Kelman, I., Chambers, J., Ayeb-Karlsson, S.: Correlating heatwaves and relative humidity with suicide (fatal intentional self-harm). Nature (2021)

[11] Page, L.A., Hajat, S., Kovats, R.S.: Relationship between daily suicide counts and temperature in england and wales. British Journal of Psychiatry **191**(2), 106–112 (Aug 2007). https://doi.org/10.1192/bjp.bp.106.031948, http://dx.doi.org/10.1192/bjp.bp.106.031948

[12] Simkovich, S., Goodman, D., Roa, C., Crocker, M., Gianella, G., Kirenga, B., Wise, R., Checkley, W.: The health and social implications of household air pollution and respiratory diseases. NPJ Prim Care Respir Med. (2019)

[13] Sing, Kaur, K., Walia: Exploring the relationship between air quality index and lung cancer mortality in india. National Library of Medicine (2023)

[14] Spirić, V.T., Janković, S., Jović, A.: The impact of air pollution on chronic respiratory diseases (2012)