IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 November 2023, accepted 8 December 2023, date of publication 14 December 2023, date of current version 21 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343149

RESEARCH ARTICLE

Automatic Detection in Twitter of Non-Traumatic Grief Due to Deaths by COVID-19: A Deep Learning Approach

JACINTO MATA-VÁZQUEZ [1], VICTORIA PACHÓN-ÁLVAREZ [1], ESTRELLA GUALDA [2], MÍRIAM ARAUJO-HERNÁNDEZ [3], AND E. BEGOÑA GARCÍA-NAVARRO [2]

[1]I2C Research Group, CITES Research Center, Information Technology Department, Universidad de Huelva, 21071 Huelva, Spain
[2]Social Studies and Social Intervention Research Group (ESEIS), Center for Research in Contemporary Thought and Innovation for Social Development (COIDESO), Universidad de Huelva, 21071 Huelva, Spain
[3]Nursing Department, Universidad de Huelva, 21071 Huelva, Spain

Corresponding author: Jacinto Mata-Vázquez (mata@uhu.es)

ABSTRACT Non-traumatic grief can be defined as, a complex process that includes emotional, physical, spiritual, social, and intellectual behaviors and responses through which individuals, families, and communities incorporate actual, anticipated, or perceived loss into their daily lives. In the age of widespread social media usage, individuals frequently share their emotions online for various reasons. This was particularly evident during the peak of the COVID-19 pandemic and its aftermath, where many social media interactions replaced or supplemented traditional farewell and mourning practices, including communications related to deaths. Recognizing messages expressing non-traumatic grief is a crucial challenge for nursing, medicine, and socio-health interventions. This awareness could assist specialists in improving early prevention and healthcare measures. In this work we present an approach to automatically detect messages (tweets) containing non-traumatic grief by means of deep learning techniques. To this end, a corpus of Spanish-language tweets has been built using a binary label to indicate the presence or absence of non-traumatic grief and has been released for use in future research. To address this challenge, multiple monolingual and multilingual language models based on pre-trained Transformer models have been fine-tuned, performing an exhaustive search to obtain the best hyperparameter values. Through this approach and employing various oversampling and undersampling techniques to mitigate the dataset's imbalance issue, the trained models reached very good results on different evaluation metrics, achieving an accuracy, AUC-ROC, and F-measure of 0.850, 0.836, and 0.827, respectively. Our results show the significance of hyperparameter selection during the learning process and demonstrate the potential of deep learning approaches for detecting non-traumatic grief messages.

INDEX TERMS Deep learning, hyperparameters, non-traumatic grief detection, transformers, Twitter.

I. INTRODUCTION

When a person dies, a structural transformation is triggered in their environment, in the social relationships of the universe of relatives, friends, and acquaintances, and they need to readjust [1]. When we refer to non-traumatic grief, we describe the natural and self-limited process of emotionally repositioning the deceased person and continuing to live, finding a reconstructed bond with them, and being able to generate other behaviors adapted to the environment without that person [2], [3]. Such restructuring has to do with how each culture understands death and with the paradigm of life that they have integrated [4].

It is vitally to describe the characteristics of the event that surrounded the death of the loved one, the levels of attachment, the relationship that the loved one had with the

The associate editor coordinating the review of this manuscript and approving it for publication was Hui Ma.

bereaved, and other specifications for coping with grief as non-traumatic one or if, on the contrary, there is a coexistence of factors that may derive to complicated or traumatic grief. The end-of-life process has been lived in isolation, without goodbyes or accompanying rituals. The critical axis here is precisely the body's materiality, the main argument being that its absence hinders the social recognition of death and the symbolic processes of mourning and grief [5]. The rituals that accompany the end-of-life ensure a process of coping with grief in a natural, healthy, and functional way that will allow a favourable evolution of the death process and will favour a process that dignifies the death and farewell of a relative, closing the circle of the lived vital process [6].

The exceptional situation that we have experienced during COVID-19 has hindered the normal development of rituals, leading to accompaniment from a distance, without contact [7]. Society has reacted to this deficit spontaneously by moving to social networks, a natural means of expression in our current society [8]. Already in the year 2000, the effectiveness of mourning practices through social networks began to be evaluated in a context where internationalization, the rhythm of life, and even the interpretation of coping had changed [9], [10], [11]. A social debate is generated about the relevance of using these technologies in moments as decisive as the death process or the grieving responses with interactions through labels, hashtags, or other elements of the language of social networks, substituting or complementing the usual farewell and mourning practices [12], [13]. The most explored networks in the analysis of interactions and the effectiveness or not of their resources in coping with grief have been Facebook and Twitter.

Twitter has the particularity of its international projection, fast access, and information delivery. A single interaction allows us to participate in any situation happening anywhere. Diverse studies [14], [15] have confirmed that not only are social networks a good way to interact, making the distances closer but that they have allowed the development of necessary processes for coping with grief, which without the support of these means, would not have been possible.

The interest of automatically identifying tweets that verbalize actions on healthy grief allows the health professional to detect early elements of protection or risk for the development of functional grief, allowing health workers to develop strategies adapted to the specific needs of the situation that COVID-19 has generated. In addition, it will favour the necessary justification for reorganizing areas of services and health structures to manage the new health demands required by the population that has experienced the loss of loved ones during the COVID-19 period. This restructuring will provide human, material, and training resources to the staff that offers accompaniment during the phases of non-traumatic grief to favour a natural resolution of grief and avoid pathological grief, generating a greater demand for health services and support, and suffering in people.

Therefore, the aim of this study is the automatic detection, using deep learning techniques, of messages containing non-traumatic grief in social network conversations. To the best of our knowledge, no previous work has addressed the automatic detection of this type of messages. While our focus is on tweets written in Spanish, the methodology described in this paper and the conclusions obtained can be applied to other languages.

A labeled dataset is a prerequisite for conducting a supervised text detection or classification task. This work deals with a binary classification task, where each document (tweet) is assigned a positive ("1") or negative ("0") label that indicates the presence or absence of non-traumatic grief messages. So, the first stage of this study involved the construction of a new labeled dataset **NTGrief** (*COVID-19-related non-traumatic grief*). The manual labeling process was carried out by experts on health sciences and sociology, and it will be described in detail in Section III. The NTGrief dataset can be accessed on Github[1] for research purposes.

In recent years, the emergence of novel techniques based on deep learning has greatly improved the effectiveness of classification tasks on structured data, images, and text. Especially, for text classification and Natural Language Processing tasks, models based on transformers [16], [17], [18] are achieving excellent results.

BERT (Bi-directional Encoder Representations from Transformers) [19] is a transformer-based language representation model, pre-trained on a vast text corpus in a self-supervised manner. It offers excellent flexibility to quickly adapt wide range of tasks such as text classification. In recent years, numerous variations of BERT have been implemented. In this work, a comparative analysis of monolingual and multilingual BERT-based models in both Spanish and English is performed. For this purpose, a fine-tuning of some of the best performing pre-trained language models in classification tasks has been carried out. Since the dataset is imbalanced, as will be described in Section IV, undersampling and oversampling approaches are proposed to achieve a better performance of the models. Furthermore, due to the variability that the predicted values can exhibit depending on the selected hyperparameter values during training, an exhaustive search for these values has been performed. Our results reveal that models trained with these strategies achieve high effectiveness in detecting messages referring to non-traumatic grief.

The following items summarize the main contributions of our work:

- A new task is suggested in the realm of detecting social media messages.
- The creation and manual labelling of a Spanish tweets Corpus (NTGrief) using a binary classification to indicate the presence or absence of non-traumatic grief

[1]https://github.com/I2C-UHU/NT-Grief

messages. The Corpus has been released for research and evaluation purposes.

- The development of deep learning models for automatically detecting non-traumatic grief in social media messages. By using various oversampling and undersampling techniques, the models achieve a high level of effectiveness, and the results are very promising. We provide a discussion of the results from computational, healthcare, and sociological perspectives.
- A comprehensive experimentation to obtain the best combination of some of the most commonly used hyperparameters for the language model training process.
- An open project on a web platform to input new messages and predict the absence or presence of non-traumatic grief.

The rest of this paper is organized as follows: in Section II, related works are discussed, delving into prior research conducted in the field. Section III describes the Corpus used in this study and the annotation process. Section IV details the different approaches used to achieve the proposed task. Results and analysis are presented in Section V. Finally, conclusions and future work are presented in Section VI.

II. RELATED WORKS

Although in the last three years, the social and health sciences have contributed to science with many investigations related to COVID-19, related to social, economic, cultural, political, educational, and other dimensions and effects, there is a significant gap regarding research that addresses the issue of non-traumatic grief and COVID-19 from a social perspective. The gap is more significant regarding research that pursues the automatic detection of non-traumatic grief in social media or is interested in natural language processing, non-traumatic grief, and social sciences, as this work addresses. Although not many studies connect natural language processing and non-traumatic grief after the deaths of COVID-19, some literature primarily focused on traumatic or complicated grief (or emotions or public opinion), and NLP is found. For example, before COVID-19, one suggestive study focused on how youth use social media in grieving [20]. Other relevant studies explored the impact of COVID-19 on the bereaved [21] or online public opinion [22] based on Weibo users or posts. Even some studies have focused on analyzing how chats or robots in the context of the COVID-19 pandemic support some populations in order to mitigate the effects of the pandemic, including coping with loss, isolation, and other problems during the pandemic [23], [24].

Other works, more numerous in the field of social sciences, focus on qualitative and mixed methods approaches employing strategies of textual codification or categorization -similar to annotation processes- [25], [26], but without pretending to automatically detect messages through deep learning techniques, or the employment of other techniques for text-based emotion detection [27]. As Julia, Davies, and Kelly underline in [28], though there is a diversity of approaches in the research that explore the use of social media for mourning, there is a primary emphasis on qualitative approaches and content analysis in the literature, which gives particular value to this work. Following Al-Garadi, Yang, and Sarker in [29], apart from the role that natural language processing has had during the pandemic, in the future to come is also needed to fight against the devastating effects of pandemics improving and take advantage of the developments in the intersection of data science and health.

To the best of our knowledge, there are no studies for the automatic detection of non-traumatic grief messages on social media. However, there are related works about sentiment analysis of messages related to COVID-19. In [30], authors conduct a meticulous exploration of COVID-19 research utilizing sophisticated data science techniques, including artificial intelligence and machine learning. Their in-depth analysis encompasses studies, comprehensive surveys of public datasets, and an insightful understanding of the challenges faced. This exhaustive review provides profound insights into the complex dynamics of COVID-19 spread and mitigation efforts. In [31], authors explores the vast data generated daily on websites and social media platforms. By employing web news mining, the study delves into data analysis and classification. The research highlights the crucial role of web news mining, particularly in social networks, enabling accurate identification of users' demographics and geographical locations. This approach, rich in statistical insights from platforms like Twitter, aids in predicting regional morbidity rates, guiding policymakers to implement targeted educational programs. Ultimately, this proactive strategy aims to reduce COVID-19 incidence and mortality in vulnerable communities. In [32], authors perform an analysis of sentiment evolution in Spanish pandemic-related tweets utilizing a fine-tuned BERT architecture. The research uncovered patterns of sentiment, prominent areas of concern related to published news, public responses and information dissemination, among other facets. In [33], authors examine the sentiments expressed by Indian citizens regarding the COVID-19 pandemic and vaccination campaign by analyzing text messages posted on the Twitter platform. They manually build a labelled dataset of tweets and emergency phone calls. To extract the emotion during the pandemic, they used pre-trained models such as DistilBERT, RoBERTa and BERT, compared their results and proposed an ensemble model to classify emotions. In [34], researchers manually labelled 10,000 tweets with the help of three annotators and classified them into ten different emotions. In [35], authors used a dataset of 226,668 tweets related to COVID-19 that was collected between December 2019 and May 2020. They proposed an ensemble deep learning model based on majority voting for sentiment analysis. From the point of view of the impact of COVID-19 on mental health, much research has also been conducted in recent years. In [36], the authors

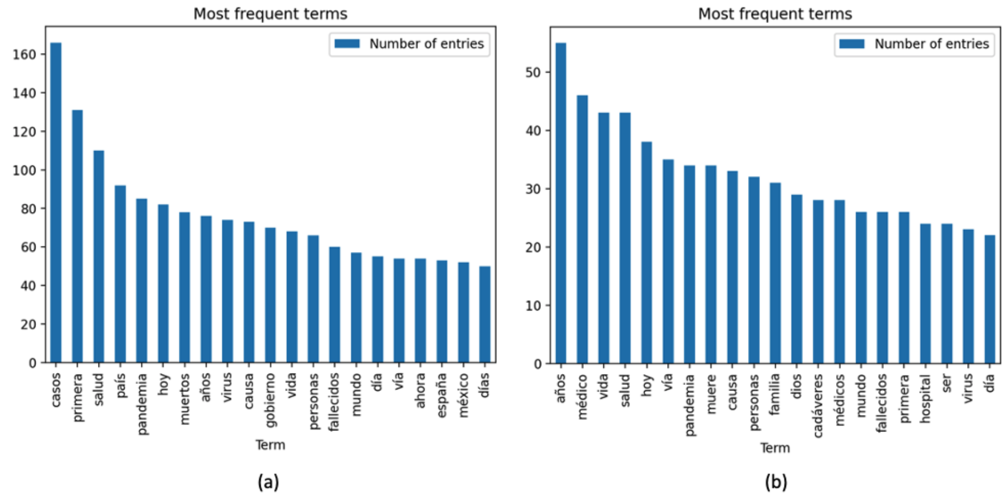

FIGURE 1. Most frequent terms (excluding search terms) used in tweets. (a) Label "0". (b) Label "1."

present a comprehensive study on the use of machine learning to combat mental disorders.

In recent years, in all work related to message detection and classification by means of natural language processing techniques in the literature, the best performance has been achieved by using language models, especially Transformer-based models.

III. CORPUS DESIGN AND CONSTRUCTION

This section introduces the NTGrief dataset, which is a Corpus designed to identify messages related to non-traumatic grief on Twitter. The process of extracting and annotating the documents is described, along with several statistics about the collection. As far as we know, NTGrief is the first Spanish Corpus developed to identify messages about non-traumatic grief in COVID-19 times.

A. DATA COLLECTION

The tweets were collected using *GetOldTweets3-optimized-modified*.[2] This is a Python project to extract old and backdated tweets. The period chosen for the searches was from January to May 2020, and a keyword-based approach was used to search for tweets. The keywords chosen for the search were "Coronavirus", "COVID", "COVID-19", "COVID_19", "COVID19" and "Muerte" ("Death"). Several searches were conducted combining each term ("Coronavirus", "COVID", "COVID-19", "COVID_19" and "COVID19") with the term "Muerte" in Spanish to collect tweets that included terms related to "Muerte & COVID-19". Since this study does not analyze dissemination patterns, retweeted messages were excluded to avoid repeated tweets. This search yielded an initial dataset of 116,936 tweets. (Death and COVID-19 dataset, 2020).

[2]https://github.com/marquisvictor/Optimized-Modified-GetOldTweets3-OMGOT

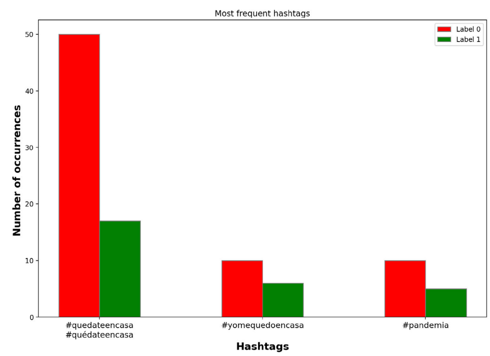

FIGURE 2. Most frequent hashtags used in tweets for both classes.

B. CORPUS ANNOTATION

To carry out a manual labeling process, it was decided to randomly obtain a significant and feasible sample to be annotated by human experts in this field. The complete methodological process followed was: (a) ten thousand tweets were randomly selected from the initial dataset. They were anonymized by replacing mentions with @USER. Additionally, emojis and URLs were removed; (b) these tweets were binary-coded by two experts and a third expert referee to solve disagreements (NTGrief_Full); (c) given that the expressions of healthy grief, in a very general context of "COVID-19 and Muerte" are a minority and this produces an very imbalanced dataset, after a first exploratory analysis, we adopted as the decision to build a more reduced dataset preserving the tweets classified as expressive of non-traumatic grief (605 tweets), and randomly removing many other tweets from the dataset to better balance the dataset. The NTGrief dataset was the outcome of this process.

As it was the first time to our knowledge that protector and risk factors of non-traumatic grief were identified and

TABLE 1. Some examples of labelled tweets in Spanish and English, along with a brief explanation of the reason for the annotation.

Spanish tweet	English version	Label	Explanation
En medio de la estela de muerte que deja el #coronavirus emerge una colonia de héroes que combaten la pandemia mientras el mundo resiste confinado para frenar la propagación. Este es un homenaje a quienes se juegan sus vidas por salvar las de otros	In the midst of the trail of death left by the #coronavirus, a colony of heroes emerges to fight the pandemic while the world resists in confinement to stop the spread. This is a tribute to those who risk their lives to save others	1	The family's tributes to the deceased protect them from traumatic grief
Después de que su esposo murió de coronavirus, ella encontró una nota de despedida emotiva en su teléfono	After her husband died of coronavirus, she found an emotional goodbye note on her phone	1	The farewell actions protect from traumatic grief
La oscuridad y la muerte no tienen la última palabra: #PapaFrancisco anima al mundo en medio del #coronavirus - Noticias Caracol vía @USER	Darkness and death do not have the last word: #PapaFrancisco animates the world in the midst of #coronavirus - Noticias Caracol via @USER	1	The manifestations of religiosity and spirituality protects from traumatic grief
Morir en tiempos de Coronavirus es realmente desaparecer. Te roban la vida. Te roban la muerte. Pienso en quienes tienen pendientes duelos y despedidas. Pienso en quienes se quedan sin el consuelo de un abrazo. Pienso en quienes se marchan solos, muy solos	Dying in times of Coronavirus is really disappearing. They steal your life. They steal your death. I think of those who have pending duels and farewells. I think of those who are left without the comfort of a hug. I think of those who leave alone, very alone	0	The lack of goodbye at the end of life promotes a traumatic grief
Fosas comunes, cuerpos en las calles, amontonados en salas... Socialismo o... muerte	Mass graves, bodies in the streets, piled up in rooms... Socialism or... death	0	The inability to see the body and not being able to perform farewell rituals promotes traumatic grief

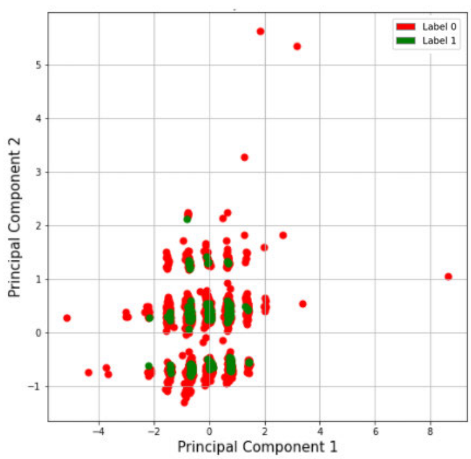

FIGURE 3. Principal Component Analysis.

classified in Twitter messages in the context of COVID-19, the process followed two steps. In a first stage, the dataset was labeled by two experts on health sciences and sociology with a first aim of distinguishing between tweets containing contents referring to protector and risk factors of non-traumatic grief, and tweets not containing that issue, even when the topics "Muerte & COVID-19" were present. Considering the task's subjective nature and level of difficulty, this dataset was manually labeled with the support of an annotation guide designed by experts in the field [37].

The overall inter-annotation agreement at this step reached a high correlation rate (0.82 of Cohen's Kappa coefficient [38]). In a second stage, those tweets in which no agreement or doubts were found were analyzed and revised carefully for previous annotators and for a third person to solve the ties if required. This in-depth revision and discussion of most difficult cases to classify helped to solve some specificities of messages published in Twitter during the pandemic and helped to discover the difficulties of identifying nuances associated with expressions of non-traumatic grief. In Table 1, several examples of annotated tweets are shown along with a brief explanation of the label chosen by the annotators.

C. DATASET FEATURES

As outlined in the previous section, an exhaustive labeling procedure of 10,000 tweets was conducted to prepare a previously categorized training dataset to carry out the supervised classification task. This labeling process resulted in an imbalanced dataset in terms of class distribution. As explained above, after a first exploratory analysis we built a more balanced dataset. Specifically, out of 2000 tweets, 1395 (69.8%) were assigned to label "0", which denotes the absence of non-traumatic grief messages, and 605 (30.2%) were assigned to label "1", indicating the presence of non-traumatic grief messages.

Text classification with imbalanced datasets is a challenging undertaking that requires data preprocessing before applying machine learning algorithms [38]. In addition, within the context of our case study, there exists minimal variance in syntax and semantics between the tweets belonging to the two distinct labels. Figure 1 shows the most frequently used terms in the tweets of the dataset for each label. On the other hand, the most frequent hashtags are the same in the texts of both labels. The top three most frequently used hashtags are shown in Figure 2.

To examine the distribution of tweets for each class in the n-dimensional vocabulary space of the dataset, a Principal Components Analysis (PCA) was carried out. Fundamentally, PCA is a technique used to reduce the dimensionality of a dataset by creating a set of new features known as Principal Components. These components are derived by transforming the original columns of the dataset. PCA offers several advantages in machine learning, including dimensionality reduction and the ability to visualize class separation, if it exists.

As can be seen in Figure 3, the classes are very overlapping. Therefore, due to the inherent complexity of the proposed task, it will be difficult to achieve optimal results.

IV. PROPOSED APPROACH

Machine learning models have been widely used to address text classification tasks. Most classical machine learning-based models extract, in a first step, some features from the documents (e.g., "Bag of Words" - BoW) and then feed them to a classifier to make a prediction. Some of these classification algorithms that have performed better on text classification tasks are Naive Bayes [39], [40], Support Vector Machines [41], [42], Hidden Markov Model [43], Gradient Boosting Trees [44] and Random Forest [45], [46]. However, models based on deep learning have recently outperformed classical machine learning approaches [47].

Transformers [48] are the most prominent architectures used for a wide range of Natural Language Processing tasks. These models are pre-trained on a large text corpus, and they can be fine-tuned on domain-specific data to achieve the best results in tasks such as text classification. BERT is a transformer-based language representation model which was designed to pre-train deep bidirectional representations from unlabeled text from BooksCorpus and English Wikipedia. In addition to pre-trained models, fine-tuning models are tuned to a specific task using a previously trained model.

In this study, we comparatively evaluated several BERT-based pre-trained language models and their expected applicability to this task.

A. LANGUAGE MODELS

This section describes the learning models with their different settings and explains the implementation decisions used in this work. The deep learning models have been developed on Google Colab[3] and they were written in Python. In particular, the Transformers library from Hugging Face [49] was employed.

There is a large number of pre-trained models available exclusively in English, so for this study, we also decided to evaluate the performance of these models using English-written tweets. For this purpose, as described in Section V, the Spanish dataset was translated into English using a machine translation tool. The following Spanish and English monolingual and multilingual transformer-based models have been used for our transfer-learning-based approach:

- **BERT**. The BERT model was proposed in [19]. BERT analyzes a sentence in two directions, that is, it considers the words that are both to the left and to the right of a keyword. This allows it to understand in depth the context and theme of the entire sentence. It was pre-trained on a large corpus of unlabeled text that includes the entirety of Wikipedia (that is, around 2.5 billion words) and the Book Corpus Dataset (800 million words). The *bert-base-uncased*[4] has been the model selected for this study.
- **BETO**. The Spanish-BERT model [50] uses a similar architecture to the BERT-Base model and was trained on a corpus that exclusively contains Spanish texts, including data from Wikipedia and the OPUS Project. Specifically, we have used the *bert-base-spanish-wwm-cased*[5] model.
- **RoBERTa**. A robustly optimized BERT pre-training approach is an improved version of BERT where key hyper-parameters are modified [51]. There are several versions of RoBERTa pre-trained in Spanish. In this work, the *roberta-base-bne*[6] [52], a model pre-trained with data from the National Library of Spain (BNE), was used for Spanish language. For English, the *roberta-base*[7] model was used.
- **XLM-RoBERTa**. It is a multilingual version of the RoBERTa model [51]. It was pre-trained on 2.5TB of a filtered CommonCrawl data containing 100 languages [53]. In this work, the *xlm-roberta-base*[8] was the selected version.
- **Multilingual-BERT (mBERT)**. It is a multilingual version of the BERT-Base model pre-trained on a corpus comprising Wikipedia texts from 104 languages [19]. The version fine-tuned in this study has been *bert-base-multilingual-cased*.[9]
- **DeBERTa**. It is a language model that aims to improve the BERT and RoBERTa models with two novel techniques: a disentangled attention mechanism and an enhanced mask decoder to replace the output softmax layer to predict the masked tokens for model pretraining [54]. The version fine-tuned for this work was *deberta-base*.[10]
- **DeBERTa v3**. It is an advanced language model that improves natural language understanding. It introduces Cross-View Training (CVT) and Relational Positional Encoding (RPE) to enhance contextual understanding and capture long-range dependencies. With a larger model size and extensive pretraining, DeBERTa v3 achieves state-of-the-art performance in various tasks [55]. In this work, the *deberta-v3-base*[11] was the selected version.

[3]https://colab.research.google.com/

[4]https://huggingface.co/bert-base-uncased
[5]https://huggingface.co/dccuchile/bert-base-spanish-wwm-cased
[6]https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne
[7]https://huggingface.co/roberta-base
[8]https://huggingface.co/xlm-roberta-base
[9]https://huggingface.co/bert-base-multilingual-cased
[10]https://huggingface.co/microsoft/deberta-base
[11]https://huggingface.co/microsoft/deberta-v3-base

TABLE 2. Hyperparameter space.

Hyperparameter	Values
Weight Decay	[0.1, 0.01, 0.001]
Maximum Length	[64, 128, 256]
Learning Rate	[2e-5, 3e-5, 5e-5]
Batch Size	[16, 32, 64]

B. BALANCING APPROACHES

As mentioned in Section IV, the NTGrief dataset is significantly imbalanced. The problem of class imbalance is a frequent challenge in machine learning classification tasks. It's necessary to implement approaches that counterbalance the imbalance to prevent the models from prioritizing the classification of the majority class. Resampling is the most widely used approach to address the issue of imbalance in text classification. Resampling is a simple yet effective solution to the class imbalance problems [56]. Random Over-Sampling (ROS) and Random Under-Sampling (RUS) are two of the most straightforward approaches to resampling. ROS will increase the amount of minority class samples, while RUS will decrease the amount of majority class samples. To overcome this imbalance problem with this technique, in this work we use a random under-sampling approach to balance the number of tweets of both classes. This approach risks losing valuable information, especially if a significant proportion of samples are removed. To test the performance of RUS, the proportion of majority and minority class samples was balanced by randomly removing majority class samples and reserving all minority class examples for training.

To carry out the oversampling, a data augmentation approach has been applied instead of using ROS. Nowadays, generating the synthesis text using data augmentation method is one way to solve imbalanced text classification problems. The method generates new samples by augmenting the existing texts using augmentation techniques such as synonym replacement, insertion, translation or deletion [57], [58], [59]. To this end, a text generation technique using *Round-Trip Translation* has been proposed to generate synthetic documents for the minority class. This technique consists of translating the tweet to a different language using an automatic translation model, and then translating it back to the target language. The use of this technique results in tweets equivalent to the original but containing different words. For this purpose, we have used the models proposed in the OPUS Project. [60]. Specifically, the models used were *opus-mt-es-de*[12] and *opus-mt-de-es*[13] to translate from Spanish to German and from German to Spanish, respectively, and *opus-mt-en-de*[14] and *opus-mt-de-en*[15] to translate from English to German and from German to English, respectively.

[12]https://huggingface.co/Helsinki-NLP/opus-mt-es-de
[13]https://huggingface.co/Helsinki-NLP/opus-mt-de-es
[14]https://huggingface.co/Helsinki-NLP/opus-mt-en-de
[15]https://huggingface.co/Helsinki-NLP/opus-mt-en-de

C. HYPERPARAMETER SEARCH

Hyperparameters are adjustable parameters that allow optimizing the training process of a pre-trained model. The performance of a model largely depends on the assignment of the most appropriate values to the hyperparameters. Table 2 shows the hyperparameter space used in the experimentation phase. For the remaining Transformers parameters, we retained the default values provided by the HuggingFace Transformers library for all the models, using GELU (*Gaussian Error Linear Unit*) as activation function for hidden layers, AdamW as the optimization function and a dropout rate of 0.1.

V. RESULTS AND ANALYSIS

In this section, we describe the experimental framework and the outcomes derived from different experiments are analyzed. Additionally, our methods of evaluation and an error analysis are described.

TABLE 3. Distribution of labels in the datasets.

Dataset	Brief name	Tweets	Label 0	Label 1
NTGrief-Full	NTGrief-F	2000	1395	605
NTGrief-Train/Validation	NTGrief-Train	1600	1116	484
NTGrief-Test	NTGrief-Test	400	279	121

TABLE 4. Automatic translation of Spanish tweets into English.

Spanish tweet	Automatic translation into English
Este tipo de noticias es para revisar lo que hemos hecho bien y mejorarlo; para corregir nuestras deficiencias; y para pensar en el modelo institucional del futuro. Costa Rica, la excepción de América durante la epidemia del coronavirus	This type of news is to review what we have done well and improve it; to correct our deficiencies; and to think about the institutional model of the future. Costa Rica, the exception of America during the epidemic of the coronavirus
La muerte se ensaña con Europa: 969 fallecidos en Italia y 796 en España por coronavirus en un día	Death is taught with Europe: 969 dead in Italy and 796 in Spain by coronavirus in one day
Perder a un ser querido siempre es muy doloroso. Mis condolencias a quienes están enfrentando la muerte de un familiar por #Covid_19 ¡Qué Dios nos ayude a todos y nos de fortaleza y sabiduría! Seguro vendrán mejores días, llenos de mucha actitud!	Losing a loved one is always very painful. My condolences to those facing the death of a relative by #Covid_19 May God help us all and give us strength and wisdom! Surely better days will come, full of much attitude!
Mi primo falleció por COVID19 el día de ayer en USA. NY. Tomemos conciencia a cualquiera le puede pasar y morir, #QuedateEnCasa	My cousin died of COVID19 yesterday in the U.S. NY. Let's make everyone aware of it and die, #StayAtHome

A. EXPERIMENTAL SETUP

First, the dataset was split, using a stratified approach, into a training dataset (80% of the data) and a test dataset (20% of the data) for the overall experimentation phase. As part of the preprocessing, all tweets were converted to lowercase, and hashtags were preserved, given our understanding that they contributed valuable semantic context during the training phase. Table 3 shows the number of tweets for each label. Thus, throughout the

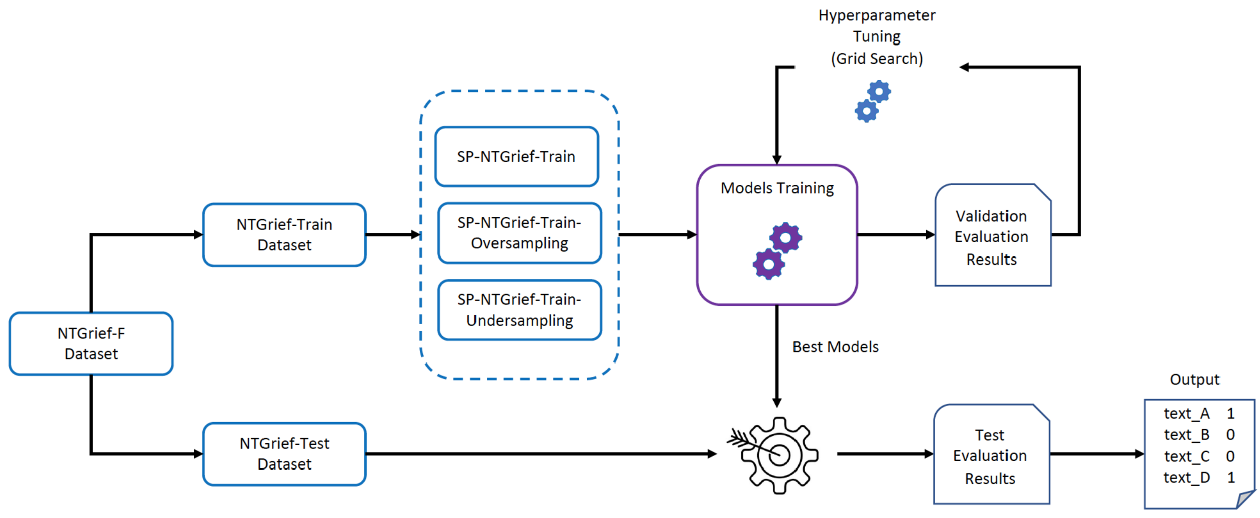

FIGURE 4. **Classification system architecture.**

TABLE 5. **Distribution of labels in the datasets after applying oversampling and undersampling techniques.**

Dataset	Tweets	Label 0	Label 1
SP-NTGrief-Train	1600	1116	484
SP-NTGrief-Train-Oversampling	2084	1116	968
SP-NTGrief-Train-Undersampling	968	484	484
SP-NTGrief-Test	400	279	121
EN-NTGrief-Train	1600	1116	484
EN-NTGrief-Train-Oversampling	2084	1116	968
EN-NTGrief-Train-Undersampling	968	484	484
EN-NTGrief-Test	400	279	121

experimentation process, the performance of all models was consistently validated using identical examples. To evaluate the performance of monolingual English pre-training models for this task, we created identical training and testing datasets that were translated into English using a machine translation form the OPUS Project [60]. In Table 4, several examples of tweets translated from Spanish to English are shown.

Finally, by applying the oversampling and undersampling techniques described above, two new datasets were constructed to train the different models selected during the experimentation phase. Additionally, to verify the effectiveness of the models in the evaluation phase, the label distribution has been maintained in the test dataset. In Table 5, the distribution of examples for each label in each of the dataset is shown.

The models were implemented using the transformers library at Hugging Face [49] and were trained on an NVIDIA Tesla T4 GPU.

A grid search was performed by trying all possible combinations of the hyperparameter space shown in Table 2. To find the optimal hyperparameter values for each model, the integration of Wandb [61] with Hugging Face was used. In the search process, epoch number was set to 10 with early stopping patience of 3 in all cases. A total of 81 runs were performed for each model and dataset, making a total of 1944 runs. Approximately 200 hours were spent on this search process.

Figure 4 shows the proposed system architecture for the detection of messages containing non-traumatic grief. In the first step, the manually labeled dataset is split into a training and a test dataset. In the next step, two new datasets are constructed using an oversampling approach and an undersampling approach. Then, an exhaustive search for the best values of the hyperparameters for the three datasets (the original and the two newly constructed) and for each selected pre-trained model is then carried out. In the next step, the models that have achieved the best results during the training and hyperparameter search process for each of the three data sets are selected. Finally, the models are evaluated using the test dataset to assess their effectiveness on unseen tweets during the training phase. The same process is carried out for both Spanish and English. The notebook including the code implemented for the experimentation is available at https://github.com/I2C-UHU/NT-Grief. Furthermore, the best performing models for English and Spanish language are available for use at https://huggingface.co/spaces/I2C-UHU/NT-Grief.

B. RESULTS

To compare the results obtained by the different models and developed strategies, a baseline based on the pre-trained Multilingual-BERT model was proposed. Given that it is not possible to know the optimal values of the hyperparameters beforehand, some of the most frequently used values were employed to perform fine-tuning of pre-trained language models: batch size of 32, learning rate of 5e-5, max length of 128 tokens and weight decay of 0.001. For this experiment, the Spanish original train dataset (SP-NTGrief-Train) was

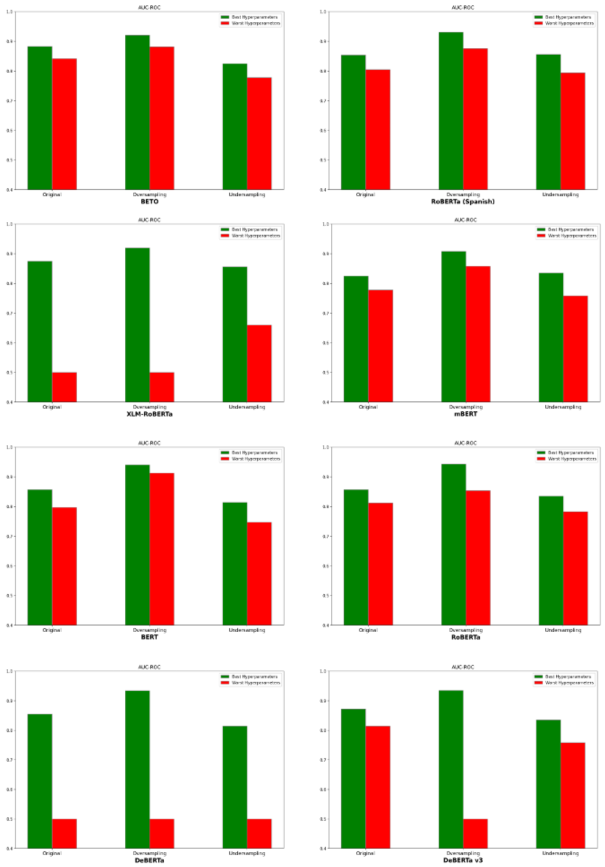

FIGURE 5. Comparative analysis of the model's performance during the training and validation stage using the best and worst values for the hyperparameters.

TABLE 6. Combination of hyperparameter values that optimize the AUC-ROC and F1-score for the minority class values for each model and dataset. Weight Decay(WD), Maximum Length(ML), Learning Rate(LR), Batch Size(BS).

Model/Dataset	WD	ML	LR	BS	AUC-ROC	F1-score Label 1
BETO						
SP-NTGrief-Train	0.1	128	5e-5	16	0.883	0.842
SP-NTGrief-Train-Over	0.01	64	5e-5	32	0.921	0.916
SP-NTGrief-Train-Under	0.1	64	5e-5	32	0.825	0.830
RoBERTa (Spanish)						
SP-NTGrief-Train	0.1	64	5e-5	16	0.854	0.783
SP-NTGrief-Train-Over	0.01	128	2e-5	64	0.931	0.926
SP-NTGrief-Train-Under	0.001	128	5e-5	64	0.856	0.860
XLM-RoBERTa						
SP-NTGrief-Train	0.001	64	2e-5	32	0.875	0.822
SP-NTGrief-Train-Over	0.1	64	3e-5	16	0.920	0.915
SP-NTGrief-Train-Under	0.01	64	3e-5	64	0.856	0.859
Multilingual-BERT (mBERT)						
SP-NTGrief-Train	0.1	64	5e-5	16	0.825	0.830
SP-NTGrief-Train-Over	0.1	64	5e-5	64	0.908	0.904
SP-NTGrief-Train-Under	0.001	64	5e-5	64	0.835	0.840
BERT						
EN-NTGrief-Train	0.01	256	5e-5	16	0.857	0.802
EN-NTGrief-Train-Over	0.001	128	5e-5	16	0.941	0.937
EN-NTGrief-Train-Under	0.01	128	5e-5	16	0.814	0.800
RoBERTa						
EN-NTGrief-Train	0.01	64	3e-5	32	0.857	0.800
EN-NTGrief-Train-Over	0.1	256	3e-5	16	0.943	0.939
EN-NTGrief-Train-Under	0.001	128	5e-5	32	0.835	0.835
DeBERTa						
EN-NTGrief-Train	0.001	256	5e-5	64	0.855	0.785
EN-NTGrief-Train-Over	0.01	256	3e-5	32	0.934	0.930
EN-NTGrief-Train-Under	0.01	256	5e-5	32	0.814	0.808
DeBERTa v3						
EN-NTGrief-Train	0.001	64	3e-5	16	0.872	0.808
EN-NTGrief-Train-Over	0.1	256	3e-5	32	0.935	0.930
EN-NTGrief-Train-Under	0.1	64	5e-5	16	0.835	0.849

used for training and the test dataset (SP-NTGrief-Test) was used to test its performance. For this study, the effectiveness has been measured using eight different metrics: (1) accuracy, (2) area under the ROC curve, (3) area under the PR curve, (4) precision, (5) recall, (6) macro F1-score, (7) F1-score for the majority class (Label 0), and (8) F1-score for the minority class (Label 1). In Table 8, the results achieved by the baseline are shown.

In Table 6, the best hyperparameter values that optimize the AUC-ROC for each model and dataset are shown. Since it is an imbalanced dataset, the F1-score of the minority class (Label 1) is also displayed. In Table 7, the combination of hyperparameter values that achieves the lowest results in terms of AUC-ROC and F1-score for the minority class is shown. Looking at the hyperparameter values in the tables, the combination of hyperparameter values that achieve the best AUC-ROC and F1-score results differs for each model and dataset. The same is true for the combination of hyperparameter values that yield the worst results. For example, for the same model (*RoBERTa-Spanish*), the optimal values for *weight decay* were different for each dataset: 0.1 for the original dataset, 0.01 for the oversampled dataset, and 0.001 for the undersampled dataset. This shows the importance of doing a preliminary search of the hyperparameter values before fine-tuning the models.

As can be seen in Figure 5, there is a highly significant difference in the AUC-ROC values based on the choice of hyperparameter values. In some cases, the model did not correctly classify any tweet from the minority class.

Once the study of hyperparameters was concluded, the models that achieved the best performance were selected for each Spanish and English dataset. Then, they were evaluated using the test dataset. It is important to state that the test dataset was constructed with the same proportion of cases as the original (69.8% for Label 0 and 30.2% Label 1) in order to evaluate the performance of the models under the same conditions.

According to Table 9, several conclusions can be drawn. For the Spanish language, the XLM-RoBERTa model performed the best in all three datasets while in English,

TABLE 7. Combination of worst performing hyperparameter values for AUC-ROC and F1-score for the minority class for each model and dataset. Weight Decay(WD), Maximum Length(ML), Learning Rate(LR), Batch Size(BS).

Model/Dataset	WD	ML	LR	BS	AUC-ROC	F1-score Label 1
BETO						
SP-NTGrief-Train	0.001	64	5e-5	32	0.842	0.781
SP-NTGrief-Train-Over	0.001	64	2e-5	64	0.882	0.876
SP-NTGrief-Train-Under	0.001	256	2e-5	64	0.778	0.788
RoBERTa (Spanish)						
SP-NTGrief-Train	0.1	64	3e-5	16	0.805	0.706
SP-NTGrief-Train-Over	0.001	64	5e-5	16	0.876	0.869
SP-NTGrief-Train-Under	0.1	64	5e-5	32	0.794	0.783
XLM-RoBERTa						
SP-NTGrief-Train	0.01	64	5e-5	32	0.500	0.000
SP-NTGrief-Train-Over	0.01	128	5e-5	16	0.500	0.000
SP-NTGrief-Train-Under	0.1	64	5e-5	16	0.660	0.694
Multilingual-BERT (mBERT)						
SP-NTGrief-Train	0.001	64	5e-5	16	0.778	0.792
SP-NTGrief-Train-Over	0.01	64	5e-5	16	0.858	0.870
SP-NTGrief-Train-Under	0.1	256	5e-5	16	0.758	0.749
BERT						
EN-NTGrief-Train	0.01	64	2e-5	16	0.797	0.719
EN-NTGrief-Train-Over	0.01	128	3e-5	32	0.913	0.908
EN-NTGrief-Train-Under	0.1	64	3e-5	64	0.747	0.749
RoBERTa						
EN-NTGrief-Train	0.001	64	5e-5	16	0.812	0.705
EN-NTGrief-Train-Over	0.01	256	3e-5	16	0.854	0.843
EN-NTGrief-Train-Under	0.01	64	5e-5	16	0.783	0.807
DeBERTa						
EN-NTGrief-Train	0.01	64	5e-5	64	0.500	0.000
EN-NTGrief-Train-Over	0.001	64	5e-5	16	0.500	0.635
EN-NTGrief-Train-Under	0.1	256	5e-5	64	0.500	0.667
DeBERTa v3						
EN-NTGrief-Train	0.1	64	5e-5	64	0.814	0.745
EN-NTGrief-Train-Over	0.1	256	5e-5	16	0.500	0.000
EN-NTGrief-Train-Under	0.01	128	2e-5	64	0.758	0.773

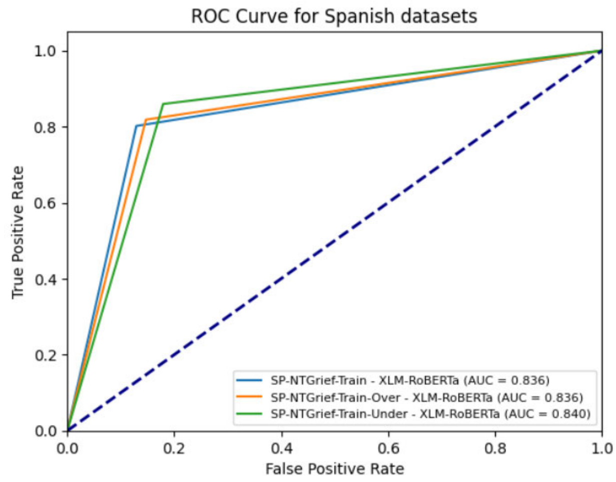

FIGURE 6. ROC Curve for the three best models on the Spanish test dataset.

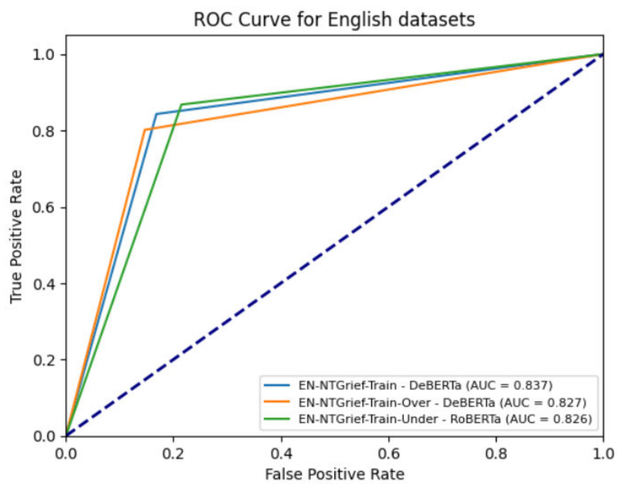

FIGURE 7. ROC Curve for the three best models on the English test dataset.

DeBERTa achieved the best results in the original dataset and in the augmented dataset, while RoBERTa was the best performing model in the reduced dataset. All methods outperform our baseline performance across all metrics. Despite the inherent difficulty of the task due to the imbalanced dataset, the results have proven to be remarkably successful. Specifically, except for the RoBERTA model, the trained models consistently achieved F1-scores above 75% for the minority class, highlighting their robustness in handling imbalanced scenarios.

In terms of the AUC-ROC and AUC-PR values, the models also exhibited very high performance, surpassing 83% and 62% respectively for the Spanish language, and comparable results for the English language. These scores highlight the models' ability to strike a favorable balance between precision and recall for both classes. Moreover, across all cases, the models consistently achieved high values in the macro F1-score measure. For a better understanding, Figures 6 and 7 show the ROC curves of the best performing models on the test dataset. As can be seen in Figure 6, the performance of the best models showcases remarkable AUC-ROC scores. The XLM-RoBERTa model exhibits adequate discriminative power across the three datasets: 0.836 on SP-NTGrief-Train, 0.836 on SP-NTGrief-Train-Over, and 0.840 on SP-NTGrief-Train-Under. These consistent high AUC-ROC values underscore the exceptional predictive accuracy of the models in message detection, demonstrating their effectiveness across diverse contexts. These findings affirm the robustness of the models and their potential for detect non-traumatic grief messages. Figure 7 shows the AUC-ROC values of the three best performing models using the English dataset. The DeBERTa model obtained the best values in EN-NTGrief-Train (0.837) and EN-NTGrief-Train-Over (0.827), while the RoBERTa model obtained the best value in the EN-NTGrief-Train-Under dataset (0.826). All models show robust AUC-ROC scores, confirming their effectiveness in detecting non-traumatic grief messages.

TABLE 8. Results for our proposed baseline on test dataset (SP-NTGrief-Test).

Model	Accuracy	AUC-ROC	AUC-PR	Precision	Recall	F1-score macro	F1-score Label 0	F1-score Label 1
Multilingual-BERT	0.787	0.726	0.549	0.750	0.726	0.736	0.853	0.619

TABLE 9. Results for non-traumatic grief detection for proposed methods on test dataset.

Dataset	Model	Accuracy	AUC-ROC	AUC-PR	Precision	Recall	F1-score macro	F1-score Label 0	F1-score Label 1
SP-NTGrief-Train	XLM-RoBERTa	0.850	0.836	0.645	0.820	0.836	0.827	0.890	0.764
SP-NTGrief-Train-Over	XLM-RoBERTa	0.842	0.836	0.633	0.811	0.836	0.821	0.883	0.759
SP-NTGrief-Train-Under	XLM-RoBERTa	0.832	0.840	0.623	0.803	0.840	0.814	0.872	0.756
EN-NTGrief-Train	DeBERTa	0.835	0.837	0.624	0.804	0.837	0.815	0.875	0.755
EN-NTGrief-Train-Over	DeBERTa	0.837	0.827	0.623	0.806	0.827	0.814	0.880	0.750
EN-NTGrief-Train-Under	RoBERTa	0.810	0.826	0.592	0.784	0.826	0.793	0.852	0.734

TABLE 10. Tweets with the manual label ('Observed') and the label predicted by the model.

#	Tweet	Observed	Predicted
1	no sabéis la alegría que siento cada vez que veo que hay menos muertos que el día anterior por causa del coronavirus #covid_19 #covid?19 #covid19esp #covid2019	1	0
2	no satanizen una enfermedad,el dar positivo a covid-19 no es sentencia de muerte, apoyemos todos a dejar la ignorancia atras y si tenemos un positivo en la familia o en los amigos tambien estarias pidiendo fotos y nombres para armar un grupo de limpieza social... seamos prudentes	1	0
3	jamás descartemos muerte como parte integral de vida. muchas más personas mueren por otras enfermedades crónicas, todos los años. mayoría no tiene ni idea son bastante más que con covid-19. pánico exagerado por redes. conocimiento es poder, economía permite buena salud.	1	0
4	no se quién pero la forma de comunicar la pandemia del coronavirus está logrando que algunas personas se mueran de miedo, vale la pena recordar que la muerte llega en el momento preciso, ni antes, ni después.....evita cambiarle la fecha	0	1
5	el jefe del gobierno local dijo que la muerte de thomas schäfer puede estar relacionada con preocupaciones debido a la crisis por la pandemia del coronavirus.	0	1

Regarding the oversampling and undersampling approaches, it can be inferred that the models exhibit similar behavior across the three datasets used for fine-tuning in each language. Our observations indicate that the performance of the models was not notably impacted by the varying number of instances available for each class in the training dataset. This finding can be attributed to the exhaustive exploration of hyperparameter values conducted for each model and training dataset. Thus, emphasizing the significance of a comprehensive hyperparameter search during the fine-tuning process of pre-trained models.

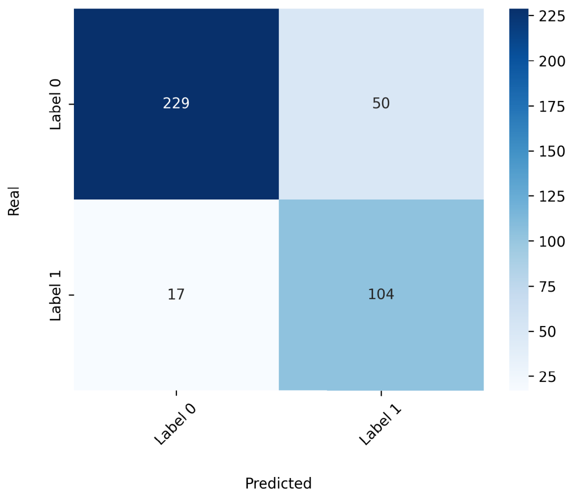

FIGURE 8. Confusion matrix for the XLM-RoBERTa model using the Spanish undersampled dataset.

C. ERROR ANALYSIS

The previous section has presented the results obtained by our models on the different metrics. Although it can be seen that a very good performance has been achieved, it is important to note that the models are not exempt from making certain prediction mistakes. To better understand the underlying causes of these failures, a comprehensive analysis of the model's mistakes was conducted. Specifically, we studied the models that best predicted the minority class for each language. For both languages, the models that best classified the minority class were those trained on the undersampling dataset. In particular, the XLM-Roberta model on the undersampling dataset managed to correctly classify 104 messages out of 121 in the minority class in the Spanish test dataset, and the RoBERTa model on the undersampling dataset correctly classified 105 messages in the English test dataset. Figures 8 and 9 show the confusion matrices of the models.

As can be seen, although the test data set is imbalanced, the models have demonstrated remarkable skill in accurately

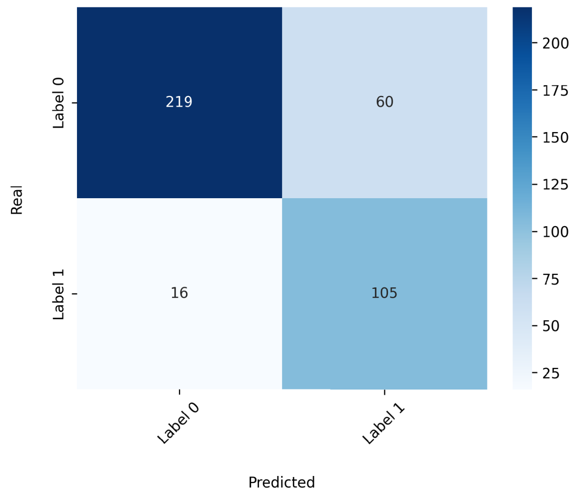

FIGURE 9. Confusion matrix for the RoBERTa model using the English undersampled dataset.

classifying the minority class. This compelling performance allows us to confidently conclude that these models are highly effective in detecting messages with non-traumatic grief content. Specifically, for this class, an 86% of accuracy is achieved for both languages. Thus, based on the evidence presented, it is a clear conclusion that the implementation of undersampling techniques has yielded remarkable results in fine-tuning the models. The application of these techniques has demonstrated to be highly effective, resulting in a significant improvement in the models' performance when it comes to classifying messages of the minority class.

For the qualitative study of error analysis, the results achieved by the Spanish model were used (Figure 8). Table 10 compiles a selection of the examples that were examined. The expert annotators manually inspected the 67 tweets that were misclassified by the model. Out of 67 entries, we identified 31 as genuine system misclassifications, while 36 where annotation errors. Of the 31 system errors, 21 were false negatives (67,8%) and 10 were false positives (32,2%).

In tweet 1 our classifier was not able to detect non-traumatic grief. We think that the reason may lie in the way the sentiment is expressed. Although the message expresses joy at the favourable evolution of the pandemic (protective factor), it is expressed in a negative way with terms such as "no sabéis" ("you don't know") or "menos muertos que el día anterior" ("fewer deaths than the day before"), leading us to think that they perceive joy at the increase in mortality due to COVID-19. In 2, the model was not able to grasp the irony and interpreted the author of the message as calling for social persecution of COVID-19 positives. However, the tweet is protecting patients who are infected from the social stigma of COVID-19. Tweet 3 addresses the concept of death as an integral part of life, providing a spiritual and value-based understanding of mortality. Giving value to death as part of life enhances the concept of transience and being mortal, thus suggesting a non-traumatic grief. We think that the model failed to interpret these concepts in the same sentence due to the presence of the double negation "never dismiss." Additionally, the model may have struggled because of the expressions "no tiene ni idea" ("has no idea") or "pánico exagerado" ("exaggerated panic"), which are terms associated with anger, rage, and frustration.

In the case of false positives, tweet 4 was manually labeled as "0" since it blames the media (press, radio, etc.) for the fear that society has towards death from the coronavirus. Expressions of blame that externalize responsibilities, as well as connotations of anger and rage, are determined as risk factors. However, expressions like "la muerte llega en el momento preciso" ("death comes at the right moment") are associated with spirituality and a sense of life's purpose. We think that may have been the reason why the model detected non-traumatic grief. Tweet 5 was manually labeled as "0" since it directly attributes the suicide of Thomas Schäfer to the pressure received during the coronavirus pandemic. However, since the accusation is not explicitly stated, the model was unable to capture that sentiment.

VI. DISCUSSION

In this paper, we conducted an in-depth analysis to understand how people cope with grief in the face of death caused by a pandemic-induced disease. Specifically, the aim is to detect messages in Twitter that denote protective factors in grief, i.e., non-traumatic grief, as opposed to messages that indicate risk factors. To this end, NT-Grief, the first Spanish corpus of expressions of grief in the face of death from COVID-19 on Twitter, was introduced. NT-Grief consists of 2000 tweets labeled as a non-traumatic grief message or not annotated by two experts and a third expert referee to solve disagreements. As far as we know, this is the first resource of this type for Spanish texts.

To address this text classification task, an exhaustive experimental study focused on natural language processing and deep learning has been carried out. Specifically, we used transfer learning from Transformers based pre-trained Large Language Models (LLM). To solve the issue of data imbalance, several sampling techniques were proposed. The experimental results showed that undersampling techniques improved the performance of the models, obtaining better results for the test dataset. On the other hand, the importance of appropriately selecting the hyperparameter values during the model training stage was demonstrated. A thoroughly study of searching for optimal hyperparameter values showed that there can be a significant difference in model performance depending on the values selected.

VII. CONCLUSION AND FUTURE WORK

We can conclude that the language models developed and presented in this work achieve outstanding performance for detecting non-traumatic grief messages in Spanish and English. The best of them achieved an accuracy of 86% in the detection of non-traumatic grief expressions, despite the large imbalance present in the training and test datasets. Furthermore, to understand the errors committed by the proposed system, an error analysis was conducted. It was detected that the errors are primarily attributed to linguistic features such as irony, ambiguity in human language, or lack of context in the messages, what poses a challenge for future works.

In a world that has been recently described as "The Platform Society" by the importance of the interactions and communications published and disseminated through different social media and their connections with societal structures [62], also considering that people frequently share their experiences, opinions, and feelings through these platforms, the advances in the automatic and quick detection of messages that express non-traumatic grief are fundamental to experts to improve early prevention and intervention in the health field. In addition, the advances made through this article will be able to explore the detection of non-traumatic grief in other datasets. Still, a step further will be to analyze the diversity of situations found as "absence of non-traumatic grief" in messages. In addition, as future work we intend to extend the scope of our models to detect various categories of non-traumatic grief messages. For this purpose, a new dataset including various categories of messages will be created. The aim of this new research is to explore the multiple expressions that people use to communicate their grief experiences.

ACKNOWLEDGMENT

The authors would like to thank the support of Pensamiento Contemporáneo e Innovación para el Desarrollo Social (COIDESO) Research Center, Tecnología, Energía y Sostenibilidad (CITES) Research Center, Estudios Sociales e Intervención Social (ESEIS) Research Group, and Ingeniería de la Información y el Conocimiento (I2C) Research Group, University of Huelva, Spain. This article is part of the Project "Approach to Death and Mourning on Social Networks (Twitter) in Times of COVID19" [Abordaje de la Muerte y Duelo en Redes Sociales (Twitter) en Tiempos de COVID19 (RSDUELO-COVID-19)] evaluated by The Andalusian Biomedical Research Ethics Portal, in July 2020 (Favorable Dictum of the Ethic Commission: 1680, n20).

REFERENCES

[1] L. Yoffe, "El duelo por la muerte de un ser querido: Creencias culturales y espirituales," *Psicodebate, Psicología, Cultura y Sociedad*, vol. 3, pp. 127–158, Jan. 2003.

[2] H. T. Herdman and S. Kamitsuru, *NANDA International Nursing Diagnoses: Definitions & Classification 2018–2020*. Thieme, 2017.

[3] M. Rodríguez-Álvaro, "Epidemiología enfermera del duelo en canarias," *Cuidar. Revista de Enfermeria de la Universidad de La Laguna*, pp. 27–48, 2020.

[4] S. M. T. Herrera and A. Acero-González, "Pérdidas y duelos complejos: Visión desde tres experiencias dolorosas vividas en Colombia," *Meridiano-Revista Colombiana de Salud Mental*, vol. 1, no. 2, pp. 13–22, Feb. 2023.

[5] G. V. L. D. S. Catela, "Mirar, desaparecer, morir. Reflexiones en torno al uso de la fotografía y los cuerpos como espacios de inscripción de la violencia," *Clepsidra. Revista Interdisciplinaria de Estudios sobre Memoria*, vol. 6, pp. 36–51, Jan. 2019.

[6] A. Soto-Rubio, E. Salvador, and M. Pérez-Marín, "Factores de riesgo de duelo complicado en cuidadores de pacientes con enfermedad terminal: Una revisión sistemática," *Calidad de Vida y Salud*, vol. 12, no. 2, pp. 29–39, 2019.

[7] M. A. Hernández, S. G. Navarro, and E. B. García-Navarro, "Abordaje del duelo y de la muerte en familiares de pacientes con COVID-19: Revisión narrativa," *Enfermería Clínica*, vol. 31, pp. 112–116, Feb. 2021.

[8] J. P. Hernández-Ramos, F. Martínez-Abad, and J. C. Sánchez-Prieto, "Empleo de videotutoriales en la era post COVID19: Valoración e influencia en la identidad docente del futuro profesional," *Revista de Educación a Distancia (RED)*, vol. 21, no. 65, Jan. 2021, Art. no. 11.

[9] S. Valero-Moreno, L. Lacomba-Trejo, A. Tamarit, M. Pérez-Marín, and I. Montoya-Castilla, "Psycho-emotional adjustment in parents of adolescents: A cross-sectional and longitudinal analysis of the impact of the COVID pandemic," *J. Pediatric Nursing*, vol. 59, pp. 44–51, Jul. 2021.

[10] S. Biçer and A. Yildirim, "Digital death and thanatechnology: New ways of thinking about data (im)mortality and digital transformation," *OMICS, A J. Integrative Biol.*, vol. 26, no. 2, pp. 88–92, Feb. 2022.

[11] E. Harrop and L. E. Selman, "Bereavement during the COVID-19 pandemic in the UK: What do we know so far?" *Bereavement J. Grief Responses Death*, vol. 1, pp. 1–9, Jan. 2022.

[12] D. McCallig, "Facebook after death: An evolving policy in a social network," *Int. J. Law Inf. Technol.*, vol. 22, no. 2, pp. 107–140, Jun. 2014.

[13] A. Burgess, "Associated technologies," in *The Executive Guide to Artificial Intelligence: How to Identify and Implement Applications for AI in Your Organization*. Springer, 2018, pp. 55–72.

[14] C. Catania-Opris, *Social Grief: A Grounded Theory of Utilizing Status Updates on Facebook as a Contemporary Ritual*. Fort Lauderdale, FL, USA: Nova Southeastern Univ., 2016.

[15] J. Moore, S. Magee, E. Gamreklidze, and J. Kowalewski, "Social media mourning: Using grounded theory to explore how people grieve on social networking sites," *OMEGA-J. Death Dying*, vol. 79, no. 3, pp. 231–259, Aug. 2019.

[16] D. RothMan, *Transformers for Natural Language Processing: Build Innovative Deep Neural Network Architectures for NLP With Python, PyTorch, TensorFlow, BERT, RoBERTa, and More*. Birmingham, U.K.: Packt, 2021.

[17] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: A review of BERT-based approaches," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5789–5829, Dec. 2021.

[18] K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, "Ammus: A survey of transformers-based pretrained models in natural language processing," *J. Biomed. Inform.*, vol. 126, 2021, Art. no. 103982.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2019, pp. 4171–4186. [Online]. Available: https://www.aclweb.org/anthology/N19-1423

[20] D. U. Patton, J. MacBeth, S. Schoenebeck, K. Shear, and K. McKeown, "Accommodating grief on Twitter: An analysis of expressions of grief among gang involved youth on Twitter using qualitative analysis and natural language processing," *Biomed. Informat. Insights*, vol. 10, Jan. 2018, Art. no. 117822261876315.

[21] N. Han, G. Chen, S. Li, F. Huang, X. Wang, X. Ren, and T. Zhu, "Impacts of the COVID-19 pandemic on the bereaved: A study of bereaved Weibo users," *Healthcare*, vol. 9, no. 6, p. 724, Jun. 2021.

[22] W. Shi, F. Zeng, A. Zhang, C. Tong, X. Shen, Z. Liu, and Z. Shi, "Online public opinion during the first epidemic wave of COVID-19 in China based on Weibo data, *Humanities Social Sci. Commun.*, vol. 9, p. 159, Dec. 2022.

[23] C. Getson and G. Nejat, "Socially assistive robots helping older adults through the pandemic and life after COVID-19," *Robotics*, vol. 10, no. 3, p. 106, Sep. 2021.

[24] A. D. Furlan, D. Zoha, B. Linor, C. Hance, B. J. Bogden, and D. Z. Buchman, "An analysis of the #CovidPain tweet chat during the first wave of the COVID-19 pandemic in 2020," *Cureus*, vol. 13, no. 10, pp. 1–12, 2021, Art. no. e18871.

[25] S. S. Mortazavi, N. Shahbazi, M. Taban, A. Alimohammadi, and M. Shati, "Mourning during corona: A phenomenological study of grief experience among close relatives during COVID-19 pandemics," *OMEGA-J. Death Dying*, vol. 87, no. 4, pp. 1088–1108, 2021.

[26] D. H.-L. Goh and C. Sian Lee, "An analysis of tweets in response to the death of Michael Jackson," *Aslib Proc.*, vol. 63, no. 5, pp. 432–444, Sep. 2011.

[27] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, "AI based emotion detection for textual big data: Techniques and contribution," *Big Data Cognit. Comput.*, vol. 5, no. 3, p. 43, Sep. 2021.

[28] J. M. Spiti, E. Davies, P. McLiesh, and J. Kelly, "How social media data are being used to research the experience of mourning: A scoping review," *PLoS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0271034.

[29] M. A. Al-Garadi, Y.-C. Yang, and A. Sarker, "The role of natural language processing during the COVID-19 pandemic: Health applications, opportunities, and challenges," *Healthcare*, vol. 10, no. 11, p. 2270, Nov. 2022.

[30] S. Latif, M. Usman, S. Manzoor, W. Iqbal, J. Qadir, G. Tyson, I. Castro, A. Razi, M. N. K. Boulos, A. Weller, and J. Crowcroft, "Leveraging data science to combat COVID-19: A comprehensive review," *IEEE Trans. Artif. Intell.*, vol. 1, no. 1, pp. 85–103, Aug. 2020.

[31] V. Rahmanian and K. Jahanbin, "Using Twitter and web news mining to predict COVID-19 outbreak," *Asian Pacific J. Tropical Med.*, vol. 13, no. 8, pp. 378–380, 2020.

[32] C. H. Miranda, G. Sanchez-Torres, and D. Salcedo, "Exploring the evolution of sentiment in Spanish pandemic tweets: A data analysis based on a fine-tuned BERT architecture," *Data*, vol. 8, no. 6, p. 96, May 2023.

[33] K. Nimmi, B. Janet, A. K. Selvan, and N. Sivakumaran, "Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset," *Appl. Soft Comput.*, vol. 122, Jun. 2022, Art. no. 108842.

[34] M. Y. Kabir and S. Madria, "EMOCOV: Machine learning for emotion detection, analysis and visualization using COVID-19 tweets," *Online Social Netw. Media*, vol. 23, May 2021, Art. no. 100135.

[35] S. Malla and P. J. A. Alphonse, "COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets," *Appl. Soft Comput.*, vol. 107, Aug. 2021, Art. no. 107495.

[36] C. Ganguly, S. Nayak, and A. K. Gupta, "Mental health impact of COVID-19 and machine learning applications in combating mental disorders: A review," in *Artificial Intelligence, Machine Learning, and Mental Health in Pandemics*. New York, NY, USA: Academic, 2022, pp. 1–51.

[37] E. B. García-Navarro, M. A. Hernández, and E. Gualda, "Factores protectores y factores de riesgo en duelo sano o funcional en tiempos de COVID-19: Guía de anotación," Grupo de Investigación Estudios Sociales E Intervención Social, Universidad de Huelva, Huelva, Spain, Tech. Rep., 2020.

[38] C. Padurariu and M. E. Breaban, "Dealing with data imbalance in text classification," *Proc. Comput. Sci.*, vol. 159, pp. 736–745, Jan. 2019.

[39] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inf. Sci.*, vol. 44, no. 1, pp. 48–59, Feb. 2018.

[40] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between multinomial and Bernoulli Naïve Bayes for text classification," in *Proc. Int. Conf. Autom., Comput. Technol. Manage. (ICACTM)*, Apr. 2019, pp. 593–596.

[41] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A novel active learning method using SVM for text classification," *Int. J. Autom. Comput.*, vol. 15, no. 3, pp. 290–298, Jun. 2018.

[42] M. Zulqarnain, R. Ghazali, Y. M. Mohmad Hassim, and M. Rehan, "Text classification based on gated recurrent unit combines with support vector machine," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 10, no. 4, pp. 3734–3742, Aug. 2020.

[43] M. Kang, J. Ahn, and K. Lee, "Opinion mining using ensemble text hidden Markov models for text classification," *Exp. Syst. Appl.*, vol. 94, pp. 218–227, Mar. 2018.

[44] F. Alzamzami, M. Hoda, and A. E. Saddik, "Light gradient boosting machine for general sentiment classification on short texts: A comparative evaluation," *IEEE Access*, vol. 8, pp. 101840–101858, 2020.

[45] T. Salles, M. Gonçalves, V. Rodrigues, and L. Rocha, "Improving random forests by neighborhood projection for effective text classification," *Inf. Syst.*, vol. 77, pp. 1–21, Sep. 2018.

[46] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Human Res.*, vol. 5, no. 1, pp. 1–16, Dec. 2020.

[47] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep learning-based text classification: A comprehensive review," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, Apr. 2022.

[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[49] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, and J. Brew, "Transformers: State-of-the-art natural language processing," in *Proc. Conf. Empirical Methods Natural Lang. Process., Syst. Demonstrations*, 2020, pp. 38–45.

[50] J. Cañete, G. Chaperon, R. Fuentes, J. H. Ho, H. Kang, and J. Pérez, "Spanish pre-trained BERT model and evaluation data," in *Proc. Practical ML Developing Countries Workshop*, 2020, pp. 1–10.

[51] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.

[52] A. Gutiérrez, J. Armengol, M. Pamies, J. Llop, J. Silveira, C. P. Carrino, C. Armentano, C. Rodríguez, A. González, and M. Villegas, "Maria: Spanish language models," *Procesamiento del Lenguaje Natural*, vol. 68, pp. 39–60, Jan. 2022.

[53] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 8440–8451.

[54] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in *Proc. ICLR*, 2020, pp. 1–23.

[55] P. He, J. Gao, and W. Chen, "DeBERTav3: Improving DeBERTa using electra-style pre-training with gradient-disentangled embedding sharing," in *Proc. ICLR*, 2021, pp. 1–16.

[56] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–50, Jun. 2017.

[57] M. Bayer, M.-A. Kaufhold, and C. Reuter, "A survey on data augmentation for text classification," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–39, Jul. 2023.

[58] S. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi, T. Mitamura, and E. Hovy, "A survey of data augmentation approaches for NLP," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 968–988.

[59] B. Li, Y. Hou, and W. Che, "Data augmentation approaches in natural language processing: A survey," *AI Open*, vol. 3, pp. 71–90, Jan. 2022.

[60] J. Tiedemann and S. Thottingal, "OPUS-MT-building open translation services for the World," in *Proc. 22nd Annu. Conf. Eur. Assoc. Mach. Translation. Eur. Assoc. Mach. Transl.*, 2020.

[61] L. Biewald, "Experiment tracking with weights and biases," vol. 2, no. 5, p. 233, 2020. [Online]. Available: https://Wandb.Com

[62] J. van Dijck, T. Poell, and M. de Waal, *The Platform Society: Public Values in a Connective World*. Oxford, U.K.: Oxford Univ. Press, 2018.

JACINTO MATA-VÁZQUEZ received the M.Sc. and Ph.D. degrees in computer science from the University of Seville, Spain. He is currently a Full Professor with the University of Huelva, Spain. He has authored or coauthored in more than 40 international journals and conferences. His research interests include text and data mining, information retrieval, and different fields of natural language processing, with an emphasis in the social media and biomedical domains.

VICTORIA PACHÓN-ÁLVAREZ received the degree in computer engineering from the University of Seville and the Ph.D. degree in computer science from the University of Huelva. She is currently an Associate Professor with the Department of Information Technologies, University of Huelva. She has published several articles in high-impact journals and has participated in competitive research projects. Her research interests include knowledge acquisition from social and biomedical databases using data mining and natural language processing techniques. She has also received awards for her teaching excellence.

ESTRELLA GUALDA received the Master of Arts degree in political science and sociology and the Ph.D. degree in sociology from the Complutense University of Madrid, Spain. She is currently a Full Professor in sociology with the University of Huelva, where she is also the Director of the Social Studies and Social Intervention Research Centre. She is also a Full Member of the Ibero-American Academy of La Rabida, Huelva, Spain. She recently published in computational sociology, COVID-19, big data, disinformation, hate speech online, conspiracy theories, and sociology of migrations.

MÍRIAM ARAUJO-HERNÁNDEZ received the degree in nursing and the Ph.D. degree in health sciences from the University of Huelva with International Mention. She is currently an Interim Substitute Professor with the Department of Nursing, University of Huelva. She is the author or coauthor of more than 12 articles in impactful journals, book chapters, international conferences, and invited papers. She has participated as a researcher and collaborated in various competitive research projects. Her research interests include end-of-life care, grief and suffering, diversity care and management, and gender.

E. BEGOÑA GARCÍA-NAVARRO received the bachelor's degree in nursing from Hogeschool Zeeland, The Netherlands, and the Ph.D. degree in health sciences from the University of Huelva with International Mention. She is currently an Assistant Professor with the Department of Nursing, University of Huelva. She has published more than 30 articles in high-impact journals and has led competitive research projects. Her research interests include the accompaniment and management of diversity at the end of life, grief, and suffering. She has received research awards for her teaching excellence and is a member of the Research Ethics Committee of the Government of Andalusia.