

## Article

# PIMA: Parameter-Shared Intelligent Media Analytics Framework for Low Resource Languages

Dimitrios Zaikis <sup>\*</sup>, Nikolaos Stylianou  and Ioannis Vlahavas <sup>\*</sup>

School of Informatics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece

<sup>\*</sup> Correspondence: dimitriz@csd.auth.gr (D.Z.); vlahavas@csd.auth.gr (I.V.)

**Featured Application:** Our work aims to provide a media analytics framework for the Greek language that utilizes subjectivity similarities among the related classification tasks, with potential for application to other low-resource languages.

**Abstract:** Media analysis (MA) is an evolving area of research in the field of text mining and an important research area for intelligent media analytics. The fundamental purpose of MA is to obtain valuable insights that help to improve many different areas of business, and ultimately customer experience, through the computational treatment of opinions, sentiments, and subjectivity on mostly highly subjective text types. These texts can come from social media, the internet, and news articles with clearly defined and unique targets. Additionally, MA-related fields include emotion, irony, and hate speech detection, which are usually tackled independently from one another without leveraging the contextual similarity between them, mainly attributed to the lack of annotated datasets. In this paper, we present a unified framework to the complete intelligent media analysis, where we propose a shared parameter layer architecture with a joint learning approach that takes advantage of each separate task for the classification of sentiments, emotions, irony, and hate speech in texts. The proposed approach was evaluated on Greek expert-annotated texts from social media posts, news articles, and internet articles such as blog posts and opinion pieces. The results show that this joint classification approach improves the classification effectiveness of each task in terms of the micro-averaged F1-score.



**Citation:** Zaikis, D.; Stylianou, N.; Vlahavas, I. PIMA: Parameter-Shared Intelligent Media Analytics Framework for Low Resource Languages. *Appl. Sci.* **2023**, *13*, 3265. <https://doi.org/10.3390/app13053265>

Academic Editors: Duy-Tai Dinh and Uday Kiran RAGE

Received: 6 February 2023

Revised: 28 February 2023

Accepted: 2 March 2023

Published: 3 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** natural language processing; media analysis; low resource languages; language model; domain adaption

## 1. Introduction

Modern day-to-day life is tightly coupled with interacting with various sources of information. Such sources span from news outlets such as CNN (<https://edition.cnn.com> (accessed on 28 February 2023)) and the BBC (<https://www.bbc.com> (accessed on 28 February 2023)), to social media sites such as Facebook (<https://www.facebook.com> (accessed on 28 February 2023)) and Twitter (<https://twitter.com> (accessed on 28 February 2023)), or even online retailers such as Amazon (<https://www.amazon.com> (accessed on 28 February 2023)) and eBay (<https://www.ebay.com> (accessed on 28 February 2023)). Media analytics is tasked with utilizing the ever-growing content presented in these public forums in an intelligent manner in order to produce valuable insights for its end users [1].

Commonly, insights such as the sentiment and the emotion expressed in such forums is important information and has been the subject of extensive research [2,3]. Additionally, depending on the type of the content's source, information such as the existence of irony and hate speech are also highly sought after in the field of social media analytics [4–7]. Subjective expressions, such as those that describe an individual's feelings towards a particular topic, brand, or product, hold great value for industries that aim to gather, analyze, and present this information in a meaningful way. Such insights can be used to

drive business directions and product revisions, and in the case of hate speech and irony detection, make platforms a safer, less toxic place [8,9]. Hence, a lot of effort is being put into developing state-of-the-art approaches for this type of content analysis.

Intuitively, it becomes clear that while all these target outcomes (i.e., emotion, sentiment, irony, and hate speech) are very different in their desired intent, they are also very tightly coupled in the way that they are expressed. However, most approaches treat them as individual goals and do not investigate the existence of underlying connections, so that one target can help define the others, and vice versa, in a uniform process and create better performing intelligent systems. This can be partially attributed to the lack of annotated parallel data, which could facilitate the investigation and exploitation of such parallel interconnections within text, leading to sequential approaches following a continuous learning paradigm.

Moreover, while this type of analysis is driven by intelligent systems, they lack the support of some, if not all, of these target outcomes for many languages other than English. These languages, characterized as low-resource due to the lack of fundamental resources, such as available training data and language specific tools, lag behind their English counterpart, as they are unable to develop intelligent tools and they face existing language specific challenges [10].

In this paper, we propose an intelligent media analytics framework for the Greek language, trained to identify sentiments, emotions, irony, and hate speech simultaneously. To facilitate this research, we also describe a newly collected and manually annotated dataset with parallel annotations for all target outcomes from diverse sources. We further investigate the connection between these different subjective expressions with respect to the performance in correctly classifying content for each target outcome through a series of experimental settings and ablations studies. Finally, we present a state-of-the-art model for Greek media analytics, which demonstrates a significant performance improvement over the individual models.

The main contributions of our work are as follows:

- A novel end-to-end transformer-based multi-task model for a joint learning approach to sentiment, irony, hate speech, and emotion classifications.
- Incorporation of a BERT-based LM and domain-adapted weights for knowledge transfer and task-specific classification heads that take advantage of the task similarities.
- Release of the described dataset and final multi-task model to facilitate further research.
- A novel training approach that can be applied to other high and low resource languages.

The remainder of this paper is organized as follows. In Section 2, we provide a brief review of related work and explore the techniques used for classification in media analytics. In Section 3, we elaborate on the datasets used for the language model domain adaption (LMDA), the multi-task classification, and our proposed framework (Section 3.3) in detail. In Section 4, we present the experimental setup and results, and finally, in Section 5, we present our conclusions and the direction of future research.

## 2. Related Works

While a number of works have been presented that deal with each aspect of our proposed framework individually in the Greek language, no work has tackled all of them simultaneously. As such, we describe and group works that individually deal with sentiment, emotion, irony, and hate speech in Greek.

Sentiment analysis of Greek texts has been a prominent research subject for over a decade [11]. However, most progress occurred during the early research years of the subject and since then has seemed to stagnate, with most approaches investigating a combination of sentiment resources and machine learning models such as naive Bayes classifiers and support vector machines [12–16].

Specifically, the use of an English and Greek sentiment and emotion lexicon has been investigated in [12], while a more targeted approach to social media using entity recognition-guided sentiment analysis, along with lexicon, is presented in [17]. Twitter specific features along

with lexicons are also investigated in different contexts in [13,16], while the importance of sentiment analysis in hotel reviews is discussed in [18]. Similarly, the focus of [19] is on identifying election trends via twitter data through sentiment analysis. Ref [14] investigated sentiment analysis on different levels of document granularity, starting from the document level and reaching the entity level, while [15] investigated the use of lexicons towards the creation of hybrid embeddings. The most recent works focus on the usability of sentiment analysis towards identifying trends in social media with respect to the pandemic, rather than creating better sentiment analysis models [20,21]. One exception is [22], which aims to create a dataset for sentiment classification of the Homeric Texts. This study identifies difficulties in reviewer agreement between different classes, which is largely based on the artistic side of the texts and the difference in perceiving contexts by the annotators.

The extensive use of lexicon has also driven research to finding better approaches towards creation of Greek specific lexicons for the task of sentiment analysis and the creation of a publicly available lexicon [23]. A recent survey on methods proposed for Greek sentiment analysis identifies the importance of word representations for Greek sentiment analysis, highlighting the performance gains acquired from different language models and the limited number of available training data [11], findings also supported by an investigative work of Greek product reviews [24]. This argument is also made stronger by the finding that language has shifted over the years [25], which suggests that datasets created in different periods to the embedding resources would not be a suitable input source.

It quickly becomes apparent, both through the findings of the previously discussed works and the methodologies used, that Greek language resources are scarce. This difference in approaches becomes more prominent when comparing methodologies for Greek sentiment analysis to those used on English documents and other non low-resource languages [26]. What is more, all works have considered only a single source of data (e.g., Twitter) and cater to the specific features present in them. In this work, we aim to solve all of these issues by utilizing a significantly larger manually annotated dataset from diverse media sources (Section 3.2). Furthermore, we use an even larger corpus from the same sources to solve issues of language evolution and align the embeddings with our modern data (Section 3.1).

Emotion classification of Greek texts is a significantly under-researched area compared to sentiment analysis, mostly due to the lack of available resources. However, due to the tight relation of the two tasks, emotion classification approaches are on par with modern Greek sentiment analysis approaches. Specifically, transformer-based approaches have been evaluated in [22,27], analyzing the effectiveness of emotion classification with different source language models. The findings of both works highlight that approaches utilizing Greek BERT [28] work best, with a domain-specific fine-tuned version of the model leading to a better performance overall.

Irony detection, on the other hand, has not been directly researched in the Greek language, primarily because it has been directly linked to aggressive rhetoric and consequently associated with hate speech [29]. In contrast, hate speech has become a very important subject of research in terms of Greek textual mining because of its importance for the community. Fueled by research interest, an offensive language detection in tweets dataset (OGTD) [30] was introduced, which sparked a plethora of approaches targeted at Greek hate speech detection. What is more, the OGTD was also part of the OffensEval 2020 [31] Task 12, targeted to offensive language in a multilingual setup, sparking the interest of the global scientific community.

Notably, BERT-based approaches, either monolingual or multilingual, have seen a lot of success with differentiating factors being the training setup and the sampling techniques used to counter imbalance [32–34]. Contest winning approaches deployed more sophisticated training paradigms and model designs [35–37]. Specifically, the addition of CNN task-specific layers to the base language model prior to the prediction head was proposed in [35]. Performance gains were also observed going from a monolingual setup to a multilingual setup in which all languages were used simultaneously with a

BERT-based model, which was also previously fine-tuned on unlabeled domain-specific data [36]. The contest wining system utilized a combination of cross-lingual learning with data augmentation techniques to enhance the data with soft labels [37].

Importantly, there have been no works that attempt to predict such a diverse range of subjectivity targets, especially in non-English languages, and in turn works have focused only on either sentiment, irony, emotion, or hate speech exclusively. In closely related works, multi-task learning has been used for hate speech-targeted English systems, with hate speech-targeted approaches utilizing similar targets to boost performance [38] and a much simpler architecture design that combines emotion and sentiment for target-specific hate speech classification [39]. Similarly, multilingual approaches for predicting monolingual sentiment fail to achieve the desired performance and cannot be used as guides to form policy [40]. Furthermore, progress appears to have stagnated to the detriment of the predictive abilities, with most works focusing only on training strategies rather than model design. Lastly, there is a limited, if any, availability of annotated data sourced from different time periods, which eliminates the possibility of creating a larger corpus by combining them due to language shifting [25].

### 3. Materials and Methods

This section describes in detail both the data and the proposed framework. Specifically, we provide a detailed view of the data collection and annotation process, along with the detailed characteristics of the datasets created as a result of the process and their intended use and an in-depth overview of our proposed framework for the multi-task, multi-class classification of sentiments and emotions and the binary classification of irony and hate speech in text.

Low resource languages (LRLs), by definition, are languages that lack the required resources to successfully tackle any NLP task. Greek is a LRL, given the scarcely available resources, both in terms of datasets to tackle tasks and the tools to assist in the creation of intelligent systems. Importantly, to successfully capture both target specific characteristics as well as period specific characteristics, we require both unlabeled domain-specific data and labeled target specific data to facilitate both an unsupervised and a supervised learning phase, respectively. The unsupervised phase is used to align a BERT-based language model (domain adaption), while the supervised phase is used to train our proposed framework's architectures on all downstream tasks. Notably, our data are collected from a variety of sources, making them more diverse than benchmark corpora even of high resource languages. Specifically, our collected data are divided into three source categories:

- Internet. A collection of articles containing blog posts, news sites, and news aggregators.
- Social Media. A collection of posts from social media sites, namely Facebook, Twitter, and Instagram.
- Press. A collection of text clips from published newspapers, identical to the printed copy and newsletters.

Both datasets consist of texts obtained through keyword-based searches on web domains related to each category. The unlabeled texts were utilized for the LM alignment, while the labeled texts were utilized for the classification tasks. Furthermore, the collected texts were filtered to obtain texts published between 2018 and 2022, while for specific user mentions on social media, search terms and accounts that were tied to official and personal accounts of corporate brands were followed to capture the relevant data. A four year gap allows for a certain degree of diversity in terms of textual content and avoids large language shifts occurring within that time period.

The framework, called PIMA (Parameter-shared Intelligent Media Analytics), consists of a parameter sharing module that is comprised of a transformer-based language model, a transformer-based encoder block, and the individual classification heads. For a more comprehensive understanding, more information regarding the attention mechanism and the BERT LM can be found in the original works where these architectures were first introduced [41,42].

### 3.1. LM Alignment Dataset

The alignment (domain adaption) dataset was collected to further pre-train Greek BERT to the media analytics domain. Adapting pre-trained LMs to domain-specific data has shown improvements in downstream tasks due to the difficulties even high-parameters models have in encoding the complexity of a single domain [43].

This dataset contains texts gathered from the same online sources and in the same domain as the ones used for classification. However, texts are considered to be sentence-level only, they come from different context than the classification dataset, and are significantly larger in volume. In total, the dataset contains 1,590,409 instances, split into 25,118,855 sentences from all source categories, with a high variance mean length of 24.2 ( $\pm 28.5$ ) tokens. The dataset statistics are presented in Table 1, in which specific information about the source type of the instances of each category are provided.

**Table 1.** Alignment (domain adaption) dataset statistics.

Domain	Category			
Internet	Blogs	News sites, aggregators		
	26,507	1,178,814		
Social media	Facebook	Twitter	Instagram	
	144,813	144,530	18,798	
Press	Free press	Sports	Economics	Local
	5212	1930	38,656	31,149

### 3.2. Media Analytics Classification Dataset

The media analytics classification dataset contains textual multi-task sentence and paragraph-level annotations for sentiment, irony, hate speech, and emotion detection. The mixture of sentence- and paragraph-level annotations in the dataset facilitates learning from minimum examples that contain any or all the task specific objectives, while allowing to generalize on a bigger body of text.

The content was randomly selected from each source to gather individual articles, posts, and text clips as uniformly distributed as possible among each source, user, and topic. Non-Greek content was excluded and achieving a diversity of content for each class was prioritized in the process.

Manual annotation was carried out by three expert annotators working professionally as journalists and was divided into batches to control the process. During the initial annotation period, a set of guidelines were devised and three rounds of 100 documents from each domain were assigned to each annotator. After each round, we evaluated the agreement of the annotators on these documents with Cohen's kappa [44], as shown in Table 2, and discussed the disagreeing texts and adjusted the guidelines. The final dataset achieved a Cohen's kappa of 0.82, which was measured among the final annotations on the complete dataset after the three pilot annotation rounds, indicating a high agreement rate.

**Table 2.** Cohen's K during the initial annotation process for guideline adjustment.

Annotation Rounds	Cohen's K	Interpretation
Round 1	0.72	Substantial agreement
Round 2	0.77	Substantial agreement
Round 3	0.86	Near perfect agreement

The final guidelines contained semantic descriptions of all targets in all four subjectivity aspects, limiting false annotations in specific types of texts for all categories, and included examples to resolve ambiguity where necessary. A few characteristic examples are described in the following:



- Promotional texts in which the launch of a new product was announced were unilaterally defined as neutral sentiment, non ironic, non hate speech containing, and of no emotion.
- Stock market reports from official financial outlets were described as neutral sentiment and no emotion.
- Ironic texts usually contain hyperbole, contradictions, metaphors and/or proverbs, blunt language, language tone change (e.g., formal in an informal context), etc.

Commonly, not all guidelines were all inclusive in terms of the subjectivity aspects, as the content can vary drastically. Evidently, there would be exceptions in these guidelines and given the expert status of the annotators they had the freedom to deviate from these guidelines in such scenarios.

To resolve disagreements, we followed a majority voting strategy on texts in which at least two of the four subjectivity labels were agreed upon by all three annotators, while discarding the disagreeing instances. The final dataset contained 14579 annotated examples with all four subjectivity labels. Table 3 shows the labels and the distribution per label for each task.

The evident imbalance in label distribution, especially for the irony and hate speech labels, can be attributed to the focus on domain-related topics of the collected content. Consequently, as post-processing step to filter and alleviate the dataset imbalance was performed by taking each labels' proportion in relation to all other labels into account, effectively reducing the over-represented labels significantly, as shown in Table 3.

**Table 3.** Classification task dataset statistics before/after post-processing

Task	Labels						
Sentiments	Negative 2146/945	Positive 2721/928	Neutral 9712/928				
Irony	No 13325/2158	Yes 1254/643					
Hate speech	No 14457/2708	Yes 122/93					
Emotion	None 11589/2031	Happiness 508/155	Contempt 2170/431	Anger 203/108	Disgust 70/58	Surprise 12/7	Sadness 27/11

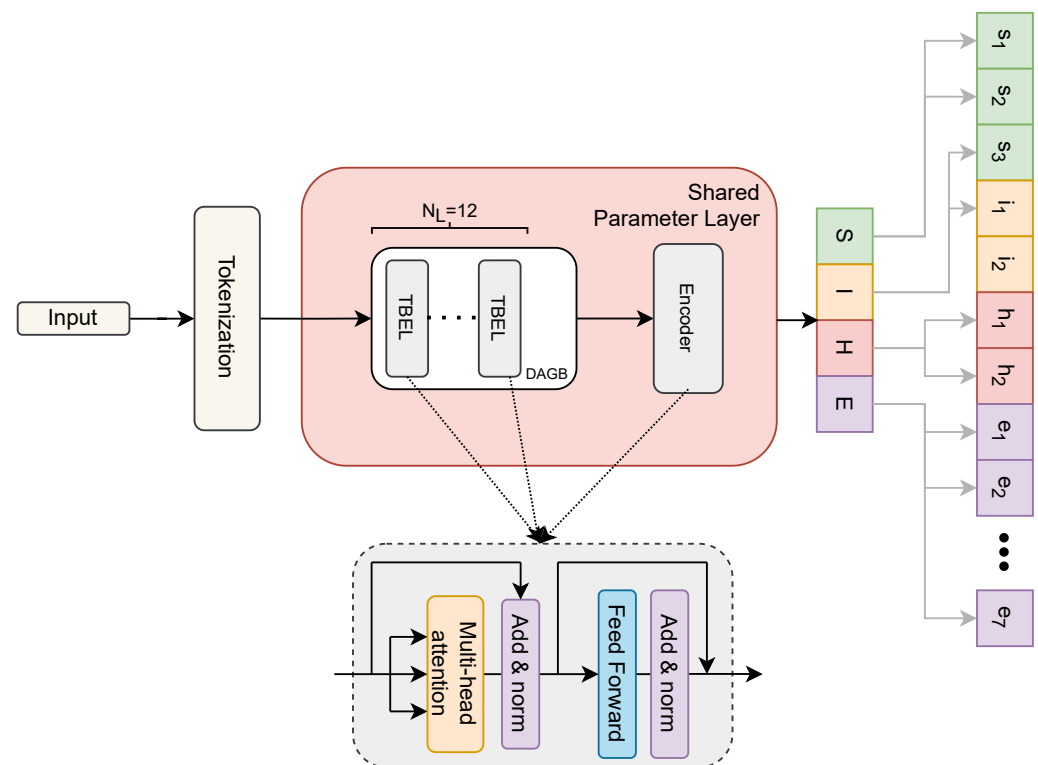
The selection of the emotion class was based on research on the basic human emotions, which was adapted for our use-case [45]. Specifically, we focused on studies performed by Ekman and Cordaro [46,47], in which these emotions are identified. In these studies, a total of seven human emotions are identified, six of which overlap with the ones used in our study, with *fear* being the one that was not utilized in our emotion classes. During our initial annotation evaluation rounds, consulting with the expert reviewers led to identifying difficulty in pinpointing characteristics in textual data that would point to fear, apart from the direct expression of the emotion using a similar word. Furthermore, given the initial sample documents, we failed to identify any evidence of such an emotion being expressed in text with none of our reviewers being able to confidently identify it. Lastly, as the original emotion studies are more holistic, i.e., not limited to text as the only medium of conveying emotion, *fear* was eliminated from the emotion classes. In comparison to these studies, we also include a "None" label for emotion, as based on our domain of application, certain types of documents are not meant to convey any emotion (i.e., reporting on events without commentary).

The final dataset used for the experiments and evaluation of our proposed framework will be made available upon publication. For each of the instances, the main body of text, the title, the related keyword, the source, and other information, such as date of publication, type of publication, and page number, were also collected.

### 3.3. Parameter-Shared Intelligent Media Analytics Framework

Our proposed framework, PIMA, is an end-to-end intelligent media analytics system for the joint classification of sentiment, emotion, irony, and hate speech in a low resource language, i.e., Greek. PIMA is comprised of a transformer-based LM and an encoder block that constitutes the parameter sharing layer and the multi-headed classification layer. The shared parameters leverage pre-trained and pre-existing knowledge from the transformer-based LM and learn multi-task representations from the joint training approach. Specifically, we use the Greek variation of BERT, named Greek BERT [28], as the baseline LM in our framework. As it was trained on modern Greek text from Wikipedia, the European Parliament Proceedings Parallel Corpus, and the Greek part of OSCAR [48], it serves as a more suitable LM for our tasks.

An overview of the framework architecture is shown in Figure 1.



**Figure 1.** Architecture overview of our proposed parameter-shared intelligent media analytics framework. S, I, H, and E denote the sentiment, irony, hate speech, and emotion classification heads, respectively, and DAGB and TBEL denote the domain-adapted Greek BERT language model and the transformer-based encoder layer, respectively.  $N_L$  denotes the number of encoder layers.

The basic architecture of the model consists of (i) the tokenized input layer, (ii) the transformer-based parameter sharing layer, and (iii) the task-specific classification heads. Initially, the model takes a sentence as a token sequence as input and captures contextual features using the pre-trained Greek BERT-based weights. An additional encoder block is applied to the initial output to generate the multi-task representations via parameter sharing.

The encoder block captures contextual information in the hidden representation layer and creates the output vector to effectively capture all relations in the input. The hidden representation is passed through the parameter sharing layer to the classification heads via linear projection. The applied task-specific feed forward layers classify the projected representations with a *Softmax* activation function.

Following this section, we first describe the parameter sharing layer and our joint end-to-end multi-task model learning approach for multi-task classification.

### 3.3.1. Parameter Sharing Layer

This layer, as its name implies, constitutes the base module of our framework, consisting of the LM and the encoder block, where the model learns shared contextual representations for all downstream tasks, namely sentiment, emotion, irony, and hate speech classification, in the context of a single input sequence.

The multi-task classification dataset contains sentences of different lengths which need to be encoded into contiguous batches of sequences. When creating these batches, padding was used to make sure that all sequences fit a given length by providing a placeholder value in the form of a “PAD” string. The LM uses a binary vector in the attention mechanism, indicating the position of these padded indices to indicate which tokens in the input sequence should be attended to.

Initially, the entire tokenized input sequence is read by the LM, generating representations while attempting to predict the masked word based on the context provided by the non-masked words in the sequence. The attention weights for each token in the input sequence, which is a combination of three embeddings (position, segment, and token embeddings), is constructed by summing them together and calculating them simultaneously. As a result, the parameter sharing layer with  $N$  multi-head self attention layers attends to each word in the sequence and attention weights are learned through parameter optimization.

The attention mechanism works by assigning weights to each word in the input sequence based on its relevance to the output. The weights determine how much attention the model should pay to each word when making predictions. In the example sentence in Table 4, the model’s attention mechanism would likely assign higher weights to certain words based on their relevance to each classification task. The words *Εξαιρετική* (excellent) and *ευχόμαστε* (hope) would likely receive higher weights as they are strong indicators of positive sentiment and emotion. The word *pantene* might also receive a relatively high weight, as it is the subject of the sentence and may provide important contextual information.

**Table 4.** Example sentences in Greek, translated to English, with the true labels.

Sentence in Greek	Πρώτη φορά βλέπω ολόκληρη τη διαφήμιση της pantene στην TV. Εξαιρετική! Δεν έγινε βέβαια από ελληνική εταιρία αλλά ευχόμαστε και οι Ελληνικές να ακολουθήσουν το παράδειγμα της.
English translation	First time I am watching the pantene commercial on TV. Excellent! Of course, it was not made by a Greek company, but we hope that Greek companies will follow its example.
Sentiment	Positive
Irony	No
Hate speech	No
Emotion	Surprise

The LM’s output is then propagated to the BERT-based encoder block that has randomly initialized weights, and thus is not pre-trained. Every linear layer in the model, including the ones inside the encoder blocks, are initialized using “normalized initialization” or more commonly called “Xavier initialization” [49], which maintains the variance of activations and back-propagated gradients all the way up or down the layers of the model. The encoder block, consisting of  $N$  encoder layers, where  $N = 1 \dots 12$ , converts the LM’s output into representations that leverage syntactic and semantic meaning by focusing on all tasks simultaneously.

### 3.3.2. Joint Multi-Task Learning

The parameters for all downstream tasks are shared through the transformer-based parameter sharing layer, with the exception of the classification heads. This approach enables the joint model to learn features and contextual information in each instance. Moreover, with the use of pre-existing knowledge in the form of the pre-trained weights



from the general and domain-specific Greek texts, the models representation learning ability is improved.

In order to minimize the effect of the individual task-specific errors, the model is trained in an end-to-end approach. However, the sentiment classification suffers during the initial training phase due to the low volume of diverse examples attributed to the classification dataset balancing. To alleviate this, the training losses are weighted to optimize the process for each task by calculating a weighted sum of the total loss, as shown in the following equation (Equation (1)),

$$L_{total} = \lambda_1 L_{sentiment} + \lambda_2 L_{irony} + \lambda_3 L_{hatespeech} + \lambda_4 L_{emotion} \quad (1)$$

where  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$  denote the sentiment, irony, hate speech, and emotion loss multipliers (LMPs), respectively.

We initially approached the joint learning task by optimizing the overall loss, where each loss has equal contribution. However, by shifting the focus to optimizing the loss multipliers in combination with the overall loss, the model's performance improved, raising the sentiment classification score significantly without sacrificing the performance of the other tasks. This approach in pre-training and multi-task focus shows an improved performance (approximately 4% averaged across all tasks) compared to equal loss contribution, since the irony and hate speech classification heads achieve good performance quickly and remain stable after reducing the  $L_{irony}$  and  $L_{hatespeech}$  loss multipliers.

Our proposed framework is designed as a multi-class classifier for sentiment, irony, hate speech, and emotion classification, where each text is assigned to a single category/class exclusively. Although the model produces probabilities for each label, the final prediction is based on the category with the highest probability value for each task.

#### 4. Results and Discussion

This section describes the experimental setting, the data preparation process, and the experimental results of our proposed framework on the four subjectivity tasks: sentiment, irony, hate speech, and emotion classification.

##### 4.1. Experimental Setting

To evaluate the performance of our framework, we conducted the experiments on the balanced dataset and used the F1-score metric, which is the harmonic mean of the precision and recall metrics. Precision and recall are the ratios of correctly predicted positive observations to the total predicted positive observations and all observations in the actual class, respectively.

We trained PIMA on all the tasks (sentiment, irony, hate speech, and emotion classification) in a joint training approach simultaneously on the balanced training examples and evaluated it on the test set, using a 80%/10%/20% split for the training, development, and test sets, respectively. The development dataset was used for the hyperparameter tuning. Furthermore, we utilized the offensive Greek Twitter dataset (OGTD) [30], which was also part of the OffensEval 2020 competition [31], as an out of distribution evaluation for our hate speech models. While both our dataset and OGTD contain Twitter posts, OGTD serves as an out of distribution dataset due to the different collection strategy and the homogeneity of the collected tweets.

We experimented with pre-trained weights from Greek BERT and our domain-adapted Greek BERT (DAGB) variant for the LM base of our parameter sharing layer. The DAGB followed the same training parameters of Greek BERT and the original BERT implementation and training objective and was trained for three epochs on the LM alignment dataset presented in Section 3.1, using AdamW [50] as an adaptive optimizer with an initial learning rate of 0.001 and a decay of 0.01 per epoch. The BERT-based architectures consisted of 12 encoder layers with a hidden layer size of 768, and input sentences were padded or truncated to the maximum sentence length. A detailed overview of the training parameters of our DAGB model are shown in Table A1 and in Appendix A.

Finally, we performed hyperparameter tuning and included the batch size, number of epochs, maximum sentence length, learning rate, dropout rate, optimizer warm-up steps, loss multipliers, and the number of encoders and encoder heads in our best parameters search objective for all experiments and models, as shown in Table 5. As the difficulty of optimization increases roughly exponentially with regard to the number of parameters in the search space, we implemented independent sampling with a pruning strategy to find the best performing combination in terms of the F1-score. Independent sampling is used for determining the value of a single parameter without considering any relationship between parameters, while the pruning strategy automatically stops unpromising trials at the early stages based on stored intermediate values. Finally, the performance evaluation of our framework and its models was tested by applying the best parameter values.

**Table 5.** Experimental parameters optimization search range

Parameter	Batch size	Epochs	Max length	Learning rate	Dropout
Range	4–64	1–20	64–512	$1 \times 10^{-6}$ – $5 \times 10^{-5}$	0.1–0.3
Parameter	warm-up steps	loss multipliers	encoders	attention heads	
Range	10–200	0–1	1–12	4–12	

LM alignment was conducted on a computer with two Nvidia RTX A6000 48GB graphics cards and a 16-core Intel CPU with an approximate training duration of 11 days. The PIMA framework was implemented with the Pytorch programming language and the experiments, which required less computing resources, were conducted on a computer with a single Nvidia RTX 3090 and a 6-core Intel CPU.

#### 4.2. Dataset Preparation

The multi-task classification dataset presented in Section 3.2 is heavily imbalanced. To balance the label distributions it was split into training, development, and testing datasets using a stratified sampling strategy to ensure the proportions of the target variables (labels) are homogeneous. This strategy inherently forces the distributions and, by extension, the imbalanced state of the dataset to remain the same. Given the application domain of our study, failing to balance the dataset would lead to heavy bias towards negative sentiment and emotions. This conclusion is supported by recent works which find that negative content spreads faster than positive content in social media [51].

Consequently, the negative class labels for the irony and hate speech tasks cover more than 90% of the examples. Moreover, the emotion task contains seven distinct labels that similarly demonstrate imbalances within themselves. In order to address this issue, a subset of the training dataset was sampled by taking each labels' proportion in relation to all other labels into account.

The official OGTD dataset, along with the training, development, and test sets, was not available during the course of this research through the official website (<https://zpitenis.com/resources/ogtd/>) (accessed on 28 February 2023). Fortunately, a version of the dataset is available through the HuggingFace dataset repository [52]. Unavoidably, both the official version and the one available to us were both imbalanced; hence, we opted to keep the unevenness between class labels.

#### 4.3. Experimental Results

In Table 6, we present our experimental results with baseline models in two settings: single target training and sequential training. As our goal is to not only provide a state-of-the-art system but also to investigate the relationship between these targets, in sequential training we continued training individual target models on the next target by only resetting the prediction head layer to match the respective target. We experimented with all possible order combinations of the tasks and present only the results of the best performing order

which is represented by Irony, Hate Speech, Emotion, and Sentiment. In all experiments presented in Table 6, we used Greek BERT in order to maintain a level of similarity with past works.

**Table 6.** Baseline experimental results

Models	Sentiment F1 (%)	Irony F1 (%)	Hate Speech F1 (%)	Emotion F1 (%)
Naive Bayes	38.18	48.56	36.20	20.30
Decision Tree	47.83	70.70	55.71	56.44
Random Forest	46.51	82.25	49.81	76.20
Logistic Regression	45.25	69.23	49.81	29.05
Support Vector Machine	27.50	48.11	49.81	12.76
Greek BERT + MLP	59.25	78.8	98.62	81.76
Greek BERT + Enc	67.76	93.17	99.54	81.99
Sentiment (IHES)	67.76	-	-	-
Irony (ESHI)	-	93.17	-	-
Hate speech (ISEH)	-	-	99.54	-
Emotion (IHSE)	-	-	-	82.15

We tested several traditional machine learning algorithms, such as naive Bayes, decision tree, random forest, logistic regression, and support vector machine, as baseline models. To align with previous works, we also implemented NN models that incorporated a language model with an MLP layer for classification (Section 2). However, our results showed that even the NN-based baseline approaches outperformed the traditional machine learning algorithms, and using a transformer-based encoder layer instead of an MLP layer significantly improved the performance across all target classes compared to the NN-based baselines. This is due to the higher-level latent representations produced by the additional layer before the prediction stage.

Our sequential models did not improve in performance in most cases, with the exception of the emotion class, when sentiment is the direct parent task. While this does indicate towards a possible connection among the target classes, this naive approach is not capable of taking advantage of the task relevance.

Table 7 details our results with our proposed joint multi-task learning approach, with parameter sharing among the target heads. We showcase results with both Greek BERT and our domain-adapted version. In the equal contribution of each target class to the model's training setup, we notice an increase in performance in irony and emotion when using our domain-adapted model, while sentiment suffers in comparison and hate speech remains constant. The performance is either equal or worse when using Greek BERT for word representations, which suggests that DAGB contributes to this performance increase.

**Table 7.** Proposed model experimental result

Models	Sentiment F1 (%)	Irony F1 (%)	Hate Speech F1 (%)	Emotion F1 (%)
PIMA				
- Greek BERT	64.1	93.1	99.54	81.3
- DAGB	65.2	93.8	99.54	82.18
PIMA /w LMP				
- Greek BERT	69.4	92.5	99.4	82.4
- DAGB	70.2	92.5	99.4	93.0

In the scaled loss (LMP) setting, however, sentiment improves dramatically, along with emotion, while irony experiences a slight performance decrease and hate speech maintains the same level of performance. Taking all scores into account, we notice that LMP-based

training yields a better overall performance, which is improved further with the use of our domain-adapted Greek BERT.

In order to further understand the experimental results, we have to take into account the label distribution in the dataset (Section 3.2). While we have made efforts to balance the data through sampling techniques (Section 4.2) which ensures scarce class examples are included, having parallel data makes it difficult to completely balance all classes. Based on this imbalance, the contribution of each class to the model's optimization needs to be adjusted, hence the performance increase in LMP. It is also due to this imbalance in the training data that sentiment experiences a performance drop when all targets are considered equally. While methods do exist to artificially generate examples [53], they work on a single class setting and would not work in our parallel corpus.

By having a parallel corpus, we were also able to effectively investigate the effects of parallel training in these four aspects of subjectivity. Our results indicate that a suspected connection between these targets exists, while the parameter sharing layer results in a performance increase compared to individual models. In two out of the three combination trained models, emotion is the final subjectivity before the target class. This result suggests a strong connection between identified emotions and both sentiment and hate speech. This is in line with findings in psychology, which highlight that the basic emotions which we use in this study are clearly separated [45]. Hence, we can deduce that clear connections between sets of emotions with different weights are directly related to sentiment and hate speech. For emotion on the other hand, the best performance is achieved with sentiment as the most recent subjectivity target. In comparison to irony and hate speech, emotion is expressed in more classes, making it easier to transfer to the six classes of emotion. It is also extremely important to note that, by definition, imbalance in the data will exist and methodologies to handle this imbalance are required in order to exploit these implicit connections.

The advantages of predicting all four targets with a single forward pass is also very important in real-world deployment of such models. Not only does it save time and computational resources, but it has the potential to increase the throughput of the system. Taking this into account, along with the overall increased performance, PIMA serves as a state-of-the-art model for Greek subjectivity classification of media content. Furthermore, while there is bound to be some information loss in the combination model between the first and the final subjectivity targets, the joint model is able to balance the information for a better overall performance, especially in sentiment and emotion target classes.

Lastly, we focus on our models' performance in an out of distribution evaluation of hate speech detection. We perform a two-fold evaluation of both NN-based baseline models and our own PIMA models, in both zero shot and fine-tuned settings, as shown in Table 8. Baseline models are composed of a BERT model, with a classification head attached for predictions, with the same hyperparameters used in our PIMA models. Importantly, the NN-based baseline zero shot models were not trained on any dataset and hence their results are the product of randomly initialized classification heads.

**Table 8.** Evaluation of our best models on OGTD dataset

Models	F1 (%)	
	Zero Shot	Fine-Tuned
NN Baseline		
- Greek BERT	20.62	33.61
- DAGB	24.65	77.32
PIMA (LMP and DAGB)		
- Hatespeech CLS	81.66	86.11
- Joint CLS	84.44	-
- Combination (ISEH)	84.44	84.44

Taking a closer look at the results, we note that our aligned (domain-adapted) Greek BERT (DAGB) performs significantly better in a fine-tuned scenario compared to Greek

BERT, while we can safely ignore the zero shot performance due to randomly initialized heads. For PIMA models, we observe a comparable performance in all three model variations during zero shot evaluation. Importantly, both the joint classification and the combination models performed equally and were improved over the directly trained hate speech model. This result is indicative of the performance benefits of parallel training in out of distribution settings. In turn, the fine-tuned hate speech model performs better due to a simpler setup, while the combination model is not able to improve in the given setting. In addition, the joint model could not be fine-tuned, as it required additional labels to compute the LMP loss for the backward propagation.

Moreover, the results on OGTD indicate of the extended usability of the proposed framework in the remaining three subjectivity targets. In zero shot, the improvements in the hate speech class can be extrapolated to sentiment, irony, and emotion classes. Unavailingly, this cannot be directly evaluated as most datasets are either dated, and hence prone to language shift [25], or not publicly available. Evidently, combination models are easier to extend to other works due to the detached nature of the subjectivity classes, while the joint model requires parallel data to be fine-tuned. However, the evidence shows no performance improvement in the combination model after being fine-tuned on the OGTD dataset.

## 5. Conclusions

In this paper, we present PIMA, an end-to-end intelligent media analytics framework for the classification of Greek text for four separate tasks: sentiment, irony, hate speech, and emotion classification. At the same time, we introduce a classification dataset that allows for the joint multi-task training on the four targets that was manually collected and annotated by experts. Our framework's model utilizes neural network parameter sharing to take advantage of the contextual similarities between the tasks in a joint fashion. We further extend our framework by domain adapting (aligning) the underlying language model to the media analytics domain with the introduction of a second dataset for this purpose.

The joint learning approach to the multi-task model in our framework leverages the individual loss functions by applying multipliers to adjust the learning process. Consequently, our proposed framework achieves a state-of-the-art performance by employing minimal feature engineering and preprocessing, coupled with parallel classifications. This approach enables high-throughput, real-world deployment and application in many media-related business areas.

In future work, we plan to extend both datasets for multi-task classification and language model adaption, making them publicly available to facilitate further research. This allows for training the underlying language model and other transformer-based LMs from scratch, potentially leading to improved NLP tools for the Greek language and low resource languages in general. Additionally, we also plan to explore other methods for improving our framework's performance by further investigating transfer learning techniques that allow us to leverage pre-trained language models from other languages. These methods have shown promising results in other domains and languages, and have the potential to significantly enhance the effectiveness and efficiency of NLP tools for low-resource languages in general.

**Author Contributions:** Conceptualization, D.Z.; methodology, D.Z. and N.S.; formal analysis, D.Z. and N.S.; investigation, D.Z. and N.S.; software, D.Z. and N.S.; resources, D.Z. and N.S.; validation, D.Z. and N.S.; writing—original draft preparation, D.Z., N.S. and I.V.; writing—review and editing, D.Z., N.S. and I.V.; visualization, D.Z.; supervision, N.S. and I.V.; project administration, D.Z. and I.V.; funding acquisition, D.Z., N.S. and I.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was carried out as part of the project KMP6-0096055 under the framework of the Action “Investment Plans of Innovation” of the Operational Program “Central Macedonia 2014-2020”, which is co-funded by the European Regional Development Fund and Greece.

**Institutional Review Board Statement:** Not applicable.



**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The complete dataset (training, development, test splits, and balanced training set), along with the code to reproduce all the experiments are available at: <https://github.com/dimitriz/pima> (accessed on 1 March 2023). The domain-adapted Greek BERT model weights are available at: <https://huggingface.co/dimitriz/greek-media-bert-base-uncased> (accessed on 1 March 2023).

**Acknowledgments:** This research was supported by the NVIDIA Applied Research Accelerator Program (ARAP) as part of the grant Advance Document Repository Mining (ADoRe).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

BERT	Bidirectional Encoder Representations from Transformers
MA	Media Analysis
NLP	Natural Language Processing
CNN	Convolutional Neural Network
MLP	Multi Layer Perceptron
LM	Language Model
LMDA	Language Model Domain Adaption
LRL	Low Resource Language
PIMA	Parameter-shared Intelligent Media Analytics framework
DAGB	Domain-Adapted Greek BERT
LMP	Loss Multipliers
OGTD	Offensive Greek Twitter Dataset

## Appendix A

**Table A1.** Training parameters for the Greek BERT-based DAGB model.

Parameter	Value
Attention layer dropout	0.1
Hidden layer dropout	0.1
Hidden layer activation function	GELU
Hidden layer size	768
Initializer range	0.02
Intermediate layer size	3072
Layer normalization eps	$1 \times 10^{-12}$
Max position embeddings	512
Number of encoder layers	12
Vocabulary size	35,000
Optimizer	AdamW
Epochs	3
Initial learning rate	0.001
Learning rate decay	0.01/epoch

## References

1. Lee, I. Big data: Dimensions, evolution, impacts, and challenges. *Bus. Horizons* **2017**, *60*, 293–303. [CrossRef]
2. Stieglitz, S.; Mirbabaie, M.; Ross, B.; Neuberger, C. Social media analytics—Challenges in topic discovery, data collection, and data preparation. *Int. J. Inf. Manag.* **2018**, *39*, 156–168. [CrossRef]
3. Manoharan, S.; Ammayappan, S. Geospatial and social media analytics for emotion analysis of theme park visitors using text mining and gis. *J. Inf. Technol.* **2020**, *2*, 100–107. [CrossRef]
4. Sykora, M.; Elayan, S.; Jackson, T.W. A qualitative analysis of sarcasm, irony and related# hashtags on Twitter. *Big Data Soc.* **2020**, *7*, 2053951720972735.
5. Potamias, R.A.; Siolas, G.; Stafylopatis, A.G. A transformer-based approach to irony and sarcasm detection. *Neural Comput. Appl.* **2020**, *32*, 17309–17320. [CrossRef]
6. Senarath, Y.; Purohit, H. Evaluating semantic feature representations to efficiently detect hate intent on social media. In Proceedings of the 2020 IEEE 14th International Conference on Semantic Computing (ICSC), San Diego, CA, USA, 3–5 February 2020; pp. 199–202.
7. Kovács, G.; Alonso, P.; Saini, R. Challenges of hate speech detection in social media. *Comput. Sci.* **2021**, *2*, 1–15. [CrossRef]

8. Giachanou, A.; Crestani, F. Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Comput. Surv.* **2016**, *49*, 1–41. [\[CrossRef\]](#)
9. Zhang, Z.; Li, X.; Chen, Y. Deciphering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews. *ACM Trans. Manage. Inf. Syst.* **2012**, *3*, 1–23. [\[CrossRef\]](#)
10. Magueresse, A.; Carles, V.; Heetderks, E. Low-resource languages: A review of past work and future challenges. *arXiv* **2020**, arXiv:2006.07264.
11. Alexandridis, G.; Varlamis, I.; Korovesis, K.; Caridakis, G.; Tsantilas, P. A survey on sentiment analysis and opinion mining in greek social media. *Information* **2021**, *12*, 331. [\[CrossRef\]](#)
12. Solakidis, G.S.; Vavliakis, K.N.; Mitkas, P.A. Multilingual sentiment analysis using emoticons and keywords. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, Poland, 11–14 August 2014, Volume 2; pp. 102–109.
13. Kalamatianos, G.; Mallis, D.; Symeonidis, S.; Arampatzis, A. Sentiment analysis of Greek tweets and hashtags using a sentiment lexicon. In Proceedings of the 19th Panhellenic Conference on Informatics, Athens, Greece, 1–3 October 2015; pp. 63–68.
14. Spatiotis, N.; Mporas, I.; Paraskevas, M.; Perikos, I. Sentiment analysis for the Greek language. In Proceedings of the 20th Pan-Hellenic Conference on Informatics, Patras, Greece, 10–12 November 2016; pp. 1–4.
15. Giatsoglou, M.; Vozalis, M.G.; Diamantaras, K.; Vakali, A.; Sarigiannidis, G.; Chatzisavvas, K.C. Sentiment analysis leveraging emotions and word embeddings. *Expert Syst. Appl.* **2017**, *69*, 214–224. [\[CrossRef\]](#)
16. Michailidis, D.; Stylianou, N.; Vlahavas, I. Real time location based sentiment analysis on twitter: The airsnt system. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence, Patras, Greece, 9–12 July 2018; pp. 1–4.
17. Petasis, G.; Spiliotopoulos, D.; Tsirakis, N.; Tsantilas, P. Sentiment analysis for reputation management: Mining the greek web. In *Proceedings of the Hellenic Conference on Artificial Intelligence*; Springer: Ioannina, Greece, 23 May 2014; pp. 327–340.
18. Markopoulos, G.; Mikros, G.; Iliadi, A.; Lontos, M. Sentiment analysis of hotel reviews in Greek: A comparison of unigram features. In *Cultural Tourism in a Digital Era*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 373–383.
19. Belevessis, D.; Tjortjis, C.; Psaradelis, D.; Nikoglou, D. A hybrid method for sentiment analysis of election related tweets. In Proceedings of the 2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNM), Piraeus, Greece, 20–22 September 2019; pp. 1–6.
20. Kydros, D.; Argyropoulou, M.; Vrana, V. A content and sentiment analysis of Greek tweets during the pandemic. *Sustainability* **2021**, *13*, 6150. [\[CrossRef\]](#)
21. Kapoteli, E.; Koukaras, P.; Tjortjis, C. Social Media Sentiment Analysis Related to COVID-19 Vaccines: Case Studies in English and Greek Language. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Crete, Greece, 17–20 June 2022; pp. 360–372.
22. Pavlopoulos, J.; Lislevand, V. Analysing the Greek Parliament Records with Emotion Classification. *arXiv* **2022**, arXiv:2205.12012
23. Tsakalidis, A.; Papadopoulos, S.; Voskaki, R.; Ioannidou, K.; Boididou, C.; Cristea, A.I.; Liakata, M.; Kompatsiaris, Y. Building and evaluating resources for sentiment analysis in the Greek language. *Lang. Resour. Eval.* **2018**, *52*, 1021–1044. [\[CrossRef\]](#)
24. Biliarios, D. Experiments in Text Classification: Analyzing the Sentiment of Electronic Product Reviews in Greek. *J. Quant. Linguist.* **2022**, *29*, 374–386. [\[CrossRef\]](#)
25. Barzokas, V.; Papagiannopoulou, E.; Tsoumakas, G. Studying the Evolution of Greek Words via Word Embeddings. In Proceedings of the 11th Hellenic Conference on Artificial Intelligence, Athens Greece, 2–4 September 2020; pp. 118–124.
26. Nandwani, P.; Verma, R. A review on sentiment analysis and emotion detection from text. *Soc. Netw. Anal. Min.* **2021**, *11*, 1–19. [\[CrossRef\]](#)
27. Alexandridis, G.; Korovesis, K.; Varlamis, I.; Tsantilas, P.; Caridakis, G. Emotion detection on Greek social media using Bidirectional Encoder Representations from Transformers. In Proceedings of the 25th Pan-Hellenic Conference on Informatics, Volos, Greece, 26–28 November 2021; pp. 28–32.
28. Koutsikakis, J.; Chalkidis, I.; Malakasiotis, P.; Androutsopoulos, I. GREEK-BERT: The Greeks visiting Sesame Street. In Proceedings of the 11th Hellenic Conference on Artificial Intelligence, Athens, Greece, 2–4 September 2020; ACM: Athens, Greece, 2020; pp. 110–117. [\[CrossRef\]](#)
29. Baider, F.; Constantinou, M. Covert hate speech: A contrastive study of Greek and Greek Cypriot online discussions with an emphasis on irony. *J. Lang. Aggress. Confl.* **2020**, *8*, 262–287. [\[CrossRef\]](#)
30. Pitenis, Z.; Zampieri, M.; Ranasinghe, T. Offensive Language Identification in Greek. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 5113–5119.
31. Zampieri, M.; Nakov, P.; Rosenthal, S.; Atanasova, P.; Karadzhov, G.; Mubarak, H.; Derczynski, L.; Pitenis, Z.; Çöltekin, c. SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020). In Proceedings of the Proceedings of SemEval, Barcelona, Spain, 12–13 December 2020.
32. Pàmies, M.; Öhman, E.; Kajava, K.; Tiedemann, J. LT@ Helsinki at SemEval-2020 Task 12: Multilingual or language-specific BERT? *arXiv* **2020**, arXiv:2008.00805.
33. Ozdemir, A.; Yeniterzi, R. SU-NLP at SemEval-2020 Task 12: Offensive Language Identification in Turkish Tweets. In Proceedings of the Fourteenth Workshop on Semantic Evaluation; International Committee for Computational Linguistics, Barcelona, Spain, 12–13 December 2020; pp. 2171–2176. [\[CrossRef\]](#)

34. Socha, K. KS@ LTH at SemEval-2020 Task 12: Fine-tuning multi-and monolingual transformer models for offensive language detection. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 2045–2053.
35. Safaya, A.; Abdullatif, M.; Yuret, D. Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 2054–2059.
36. Wang, S.; Liu, J.; Ouyang, X.; Sun, Y. Galileo at SemEval-2020 Task 12: Multi-lingual Learning for Offensive Language Identification Using Pre-trained Language Models. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 1448–1455.
37. Ahn, H.; Sun, J.; Park, C.Y.; Seo, J. NLPDove at SemEval-2020 Task 12: Improving Offensive Language Detection with Cross-lingual Transfer. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 1576–1586.
38. Kapil, P.; Ekbil, A. A deep neural network based multi-task learning approach to hate speech detection. *Knowl.-Based Syst.* **2020**, *210*, 106458. [\[CrossRef\]](#)
39. Plaza-del Arco, F.M.; Halat, S.; Padó, S.; Klinger, R. Multi-task learning with sentiment, emotion, and target detection to recognize hate speech and offensive language. *arXiv* **2021**, arXiv:2109.10255.
40. Manias, G.; Kiourtis, A.; Mavrogiorgou, A.; Kyriazis, D. Multilingual Sentiment Analysis on Twitter Data Towards Enhanced Policy Making. In Proceedings of the Artificial Intelligence Applications and Innovations: 18th IFIP WG 12.5 International Conference, AIAI 2022, Hersonissos, Crete, Greece, 17–20 June 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 325–337.
41. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017, Volume 30.
42. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Long and Short Papers; Association for Computational Linguistics: Minneapolis, Minnesota, 2019; Volume 1; pp. 4171–4186. [\[CrossRef\]](#)
43. Gururangan, S.; Marasović, A.; Swayamdipta, S.; Lo, K.; Beltagy, I.; Downey, D.; Smith, N.A. Do not Stop Pretraining: Adapt Language Models to Domains and Tasks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020.
44. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [\[CrossRef\]](#)
45. Tracy, J.L.; Randles, D. Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emot. Rev.* **2011**, *3*, 397–405. [\[CrossRef\]](#)
46. Ekman, P. Facial expressions of emotion: New findings, new questions. *Psychol. Sci.* **1992**, *3*, 34–38. [\[CrossRef\]](#)
47. Ekman, P.; Cordaro, D. What is meant by calling emotions basic. *Emot. Rev.* **2011**, *3*, 364–370. [\[CrossRef\]](#)
48. Suarez, P.J.O.; Sagot, B.; Romary, L. Asynchronous pipelines for processing huge corpora on medium to low resource infrastructures; Leibniz-Institut für Deutsche Sprache: Mannheim, 2019. In Proceedings of the Workshop on Challenges in the Management of Large Corpora (CMLC-7) 2019, Cardiff, UK, 22 July 2019; pp. 9–16. [\[CrossRef\]](#)
49. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.S.; Asari, V.K. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* **2019**, *8*, 292. [\[CrossRef\]](#)
50. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the ICLR, New Orleans, LA, USA, 6–9 May 2019.
51. Antypas, D.; Preece, A.; Camacho-Collados, J. Negativity spreads faster: A large-scale multilingual twitter analysis on the role of sentiment in political communication. *Online Soc. Netw. Media* **2023**, *33*, 100242. [\[CrossRef\]](#)
52. Lhoest, Q.; del Moral, A.V.; Jernite, Y.; Thakur, A.; von Platen, P.; Patil, S.; Chaumond, J.; Drame, M.; Plu, J.; Tunstall, L.; et al. Datasets: A community library for natural language processing. *arXiv* **2021**, arXiv:2109.02846.
53. Bansal, M.A.; Sharma, D.R.; Kathuria, D.M. A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications. *Acm Comput. Surv. (Csur)* **2021**, *54*, 1–29. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.