mathematics

Article

Influenza-like Illness Detection from Arabic Facebook Posts Based on Sentiment Analysis and 1D Convolutional Neural Network

Abdennour Boulesnane [1,*], Souham Meshoul [2,*] and Khaoula Aouissi [3]

[1] BIOSTIM Laboratory, Medicine Faculty, Salah Boubnider University Constantine 03, Constantine 25001, Algeria
[2] Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[3] Department of Pharmacy, Medicine Faculty, Salah Boubnider University Constantine 03, Constantine 25001, Algeria
* Correspondence: aboulesnane@univ-constantine3.dz (A.B.); sbmeshoul@pnu.edu.sa (S.M.)

Abstract: The recent large outbreak of infectious diseases, such as influenza-like illnesses and COVID-19, has resulted in a flood of health-related posts on the Internet in general and on social media in particular, in a wide range of languages and dialects around the world. The obvious relationship between the number of infectious disease cases and the number of social media posts prompted us to consider how we can leverage such health-related content to detect the emergence of diseases, particularly influenza-like illnesses, and foster disease surveillance systems. We used Algerian Arabic posts as a case study in our research. From data collection to content classification, a complete workflow was implemented. The main contributions of this work are the creation of a large corpus of Arabic Facebook posts based on Algerian dialect and the proposal of a new classification model based on sentiment analysis and one-dimensional convolutional neural networks. The proposed model categorizes Facebook posts based on the users' feelings. To counteract data imbalance, two techniques have been considered, namely, SMOTE and random oversampling (ROS). Using a 5-fold cross-validation, the proposed model outperformed other baseline and state-of-the-art models such as SVM, LSTM, GRU, and BiLTSM in terms of several performance metrics.

Keywords: influenza-like illness; COVID-19; Arabic sentiment analysis; disease classification; Facebook; Algerian dialect

MSC: 68T07

check for updates

Citation: Boulesnane, A.; Meshoul, S.; Aouissi, K. Influenza-like Illness Detection from Arabic Facebook Posts Based on Sentiment Analysis and 1D Convolutional Neural Network. *Mathematics* **2022**, *10*, 4089. https://doi.org/10.3390/math10214089

Academic Editors: Carmen Lacave and Ana Isabel Molina

Received: 9 October 2022
Accepted: 31 October 2022
Published: 2 November 2022

1. Introduction

The huge increase in the use of social media platforms has made them an important source of massive amounts of data. Users of social media are now sharing every aspect of their lives, including their political beliefs, emotional feelings, health status, anxiety, anger, and even their wishes. Based on Social Media Analysis (SMA) [1], such data have been used for a variety of purposes, including product marketing [2], political elections [3], tourism [4], healthcare [5], and renewable energy [6,7], among others.

With over 2.9 billion users [8], Facebook is one of the world's largest social networking platforms, allowing the sharing of diverse data in a variety of daily life domains. Algeria has an estimated 27 million Internet users, accounting for 60% of the total population. About 22.4 million of these people use Facebook. As a result, Facebook is the most popular social media platform in this country [9].

As is the case in the rest of the world, Algerian Facebook users have recently and extensively shared a great deal of health-related information, including requests for medical advice and fears of certain diseases, especially in light of the rising incidence of rapidly spreading infectious diseases such as Influenza-Like Illnesses (ILI) and COVID-19. The Centers for Disease Control and Prevention [10] define ILI as "a fever, cough, and/or sore throat with no other known cause than influenza". While COVID-19 can have severe consequences and cause organ damage, its clinical manifestations are comparable to those of the common cold, such as fever, cough, and sore throat [11,12].

On the other hand, health systems still rely significantly on health center data to detect diseases and follow their spread, which is a time-consuming and labor-intensive process prior to issuing public warnings. Therefore, it has become imperative to strengthen existing health systems by leveraging health-related data on social media and developing intelligent systems that help in monitoring the spread of infectious diseases such as ILI, anticipating and controlling outbreaks, providing early warnings, and identifying the emergence of new symptoms.

Several studies have been undertaken to improve public health systems by leveraging social media health-related data, machine or deep learning models, and Natural Language Processing (NLP) techniques, such as text mining and sentiment analysis. These studies include the detection of various diseases through social networks, such as COVID-19 [13–15], latent infectious diseases [16], infectious diseases [17], depression [18–20], mental illness [21,22], mosquito-borne diseases [23], Asperger syndrome [24], dengue disease [25], avian influenza [26], and influenza [27–31], among others.

However, as these works rely on an in-depth comprehension of the natural language used to analyze emotions and detect diseases from published texts, their use is mostly limited to this language, and they cannot be used for other natural languages. Moreover, to the best of our knowledge, no previous research has used sentiment analysis on social media data written in the Algerian Arabic dialect to detect diseases.

In this paper, we present a new sentiment classification model based on one-dimensional convolutional neural networks (1D-CNN) and sentiment analysis to detect and monitor ILI in Facebook postings from Algeria. The suggested approach is able to interpret the emotions of Algerian-speaking patients and identify ILI-positive instances. This work's contributions can be summarized as follows: (1) A corpus of 21,885 Facebook posts written in Arabic Algerian dialect was compiled. This data set comprises health-related information that can be utilized by a variety of medical applications for the benefit of the public health. (2) All acquired data were manually annotated by professionals, enabling the development of a model capable of comprehending how a patient with ILI is feeling. (3) We examined, balanced, and preprocessed the data as part of the data preparation phase by implementing novel NLP approaches, such as recommending new stop words appropriate to the Algerian Arabic dialect. (4) Multiple Feature Extraction (FE) approaches were employed, and a methodology called "Feature concatenation" was introduced to improve the extraction process by merging these methods. (5) We propose a new 1D-CNN-based model architecture with many layers trained to identify and classify ILI from Facebook postings. Finally, an extensive evaluation process was undertaken to show the effectiveness of the proposed approach.

The remainder of the paper is organized as follows: Section 2 discusses the most recent works on Arabic sentiment analysis related to public health. Section 3 describes the proposed approach in detail. The results of the experiments are discussed and analyzed in Section 4. Section 5 includes a conclusion and presents future work plans.

2. Background and Related Work

Compared to other languages such as English, Spanish, and Chinese, Arabic remains considerably less prevalent on the Internet. Moreover, for the purposes of NLP, Arabic content requires significantly more effort to extract the sentiment and core idea behind the text, as nearly every Arabic-speaking nation utilizes a different dialect. Furthermore, regarding Arabic health-related content on social media, it is not being used effectively to benefit public health on the one hand, and on the other hand, users lack the awareness required to safeguard their sensitive data [32].

In this section, we will present an overview of recent works in the literature that apply sentiment analysis techniques [33] based on deep or machine learning and use social media health-related data written in the Arabic language and/or its dialects.

In [34,35], sentiment analysis using Machine Learning (ML) was adopted to understand and analyze the social behavior of Saudi individuals towards certain health services (such as mHealth apps) and to assess the extent of their awareness of the quarantine during the COVID-19 pandemic. Each study collects, labels, processes, and sentimentally classifies Arabic tweets into three categories, namely, "positive", "negative", and "neutral".

An Arabic language dialect identification system is proposed in [36], aiming to analyze and classify COVID-19-related tweets into four Arabic dialects: Modern Standard Arabic (MSA), Egyptian, Gulf, and Levantine. In this study, BERT-based models were adopted to locate the source region of COVID-19 Arabic tweets, thus helping to monitor the epidemic outbreaks in the Arab world. Furthermore, the data from [37,38] were used, and the features were extracted based on Term Frequency-Inverse Document Frequency (TF-IDF) and word embedding. As a result, the proposed system achieved a very strong performance in determining the tweets' sources with an accuracy of 97.36%.

In [39], COVID-19 vaccine-related tweets were collected and analyzed for six Gulf countries to study people's feelings about different types of vaccines to support the vaccination process. The collected data were cleaned, tokenized, and then scored using three sentiment analysis methods, TextBlob, Ratio, and VADER, producing positive and negative instances. After that, the LSTM was used to extract deep features and provide them to ML classifiers, including SVM, Fine-KNN, and Ensemble Boost. The best sentiment classification results were achieved for fine-KNN and Ensemble boost classifiers with accuracy of 94.01%.

In [40], more than 4.5 million Arabic tweets were collected related to the topic of COVID-19. The main objective of this study was to detect rumors and misinformation about COVID-19 in Arabic content. For this purpose, 8786 tweets were annotated into two categories—"misinformation" and "not", based on a list of misinformation collected from reliable sources. Furthermore, using TF-IDF and other word embedding methods such as word2veca and FASTTEXT, the features were extracted and then fed to several ML and deep learning models.

In another similar study [41], an AraBERT-based model was proposed that can determine whether Arabic health-related tweets are accurate or not. This work focuses on training and evaluating the performance of various deep learning models that use transformer models and pretrained word embeddings. The results demonstrated the efficacy of the AraBERT-based model over the other deep learning models in identifying the medical accuracy of Arabic tweets.

In [42], Arab tweets were used to build a monitoring system to track and analyze people's emotions during the spread of COVID-19, as well as to monitor the symptoms that appear as a result this disease. Using rule-based (if-then) techniques, 5.5 million tweets were collected and annotated for their study. Additionally, two types of classification were adopted, namely, emotion-based multi-class classification and symptom-based binary classification. Initially, the LSTM deep learning model is used to classify Arabic tweets into six emotions, including "anger", "disgust", "fear", "joy", "sadness", and "surprise". Then, a second LSTM classifier is introduced to classify tweets into either "symptom" or "non-symptom" categories.

Another similar study [43] intends to build a health monitoring system in order to discover concerns associated with the COVID-19 epidemic and to assess the sentiments of Moroccan users on Facebook, Twitter, YouTube, and other popular websites. In addition to the Arabic language, the researchers focused on the Moroccan dialect and developed MD-ULM, the first Universal Language Model for the Moroccan dialect. This proposed model is mainly based on LSTM to classify text comments by topic and emotion.

Two BERT-based models for analyzing Arabic tweets and evaluating the influence of COVID-19 on users' mental health were proposed in [44]. In this paper, the authors propose a new method called dynamically weighted loss function to address the issue of unbalanced data. Word and contextual embeddings were used to extract features from tweets, and emojis were substituted with more expressive ones in terms of sentiment and emotion. On the basis of these methodologies, BERT-based transformers were utilized to detect sentiment in Arabic COVID-19 tweets, thereby protecting individuals from mental diseases such as depression, anxiety, and so on.

In [45], several ML models, including Random Forest (RF), AdaBoostM1, Naïve Bayes (NB), and Liblinear, were used to determine whether Twitter users in the Arab Gulf region were suffering from depression. Based on sentiment analysis and NLP, each tweet was categorized as either "Depressed" or "non-depressed". In addition to tweets written in MSA, the authors of this work also considered Arabian Gulf languages to train ML classifiers and produce more accurate models.

A similar study was conducted to aid in the diagnosis of depression in [46]. After collecting and thoroughly analyzing 4542 tweets based on nine depression symptoms, the tweets were classified into three broad sentiment categories: "non-depressed", "depressed", and "neutral". In their research, the authors extract data features from processed Arabic tweets using N-grams and TF-IDF techniques. These features were then fed into several classifiers based on ML.

On the other hand, the authors of [47] used sentiment analysis to cluster and categorize depression levels and causes accordingly. Facebook groups were used as the data source to detect and evaluate depression among Egyptian women. In addition, a cluster LSTM model was presented to determine the sex and depression levels of Facebook users based on their text comments. Furthermore, Word2vec and LSTM were employed to classify each comment into a variety of causes of depression, such as family issues, education, employment problems, sicknesses, newborns, etc.

Another interesting study [48] used YouTube comments to protect people with diabetes from misinformation by analyzing sentiments in the comments for herbal treatment videos. For this purpose, a newly compiled dataset of 4111 comments called ADHTD was developed. This dataset was split into positive and negative classes based on the annotators' analysis. Furthermore, the Synthetic Minority Oversampling Technique (SMOTE) was employed to address the uneven distribution of the ADHTD dataset. Upon this basis, the suggested ML classifiers, particularly Support Vector Machine (SVM) and Logistic Regression (LR) models, achieved great performance with up to 92% accuracy.

In [49], sentiment analysis was used to monitor influenza epidemics in tweets from Arab countries. In their work, several ML models were proposed to classify Arabic tweets into two different classes: A valid class representing influenza-related tweets and an invalid class for tweets unrelated to influenza. Although the proposed models in this study demonstrated promising results for interpreting Arabic tweets, they did not account for the diverse Arabic dialects spoken in other Arab countries. Moreover, Twitter is less popular in the Maghreb than in the Middle East.

In [50], a significant study was discussed that concerns detecting rumors and misinformation about cancer treatment spread in Arabic content on social media. In this regard, a corpus of Arabic tweets was collected and annotated into two classes: "Rumor" and "non-Rumor". As in many studies, data were processed, and features were extracted using TF-IDF. After that, several models were proposed using several feature extraction methods, with and without oversampling techniques.

Table 1 provides a brief summary of the above-discussed works related to Arabic sentiment analysis in public health. As can be seen, various text data representations and ML models were employed. The proposed models are closely related to the used language/dialect. There has been no research into the Algerian spoken dialect related to health-based content to the best of our knowledge.

Table 1. A summary of related work for recent Arabic sentiment analysis related to public health.

Articles	Model	Disease	Social Network	#Instances	#Classes	Result
[34]	SVM with AraVec Embeddings	COVID-19	Twitter	4719	3	85.00% F1
[35]	SVM with Bigram in TF-IDF	COVID-19	Twitter	242,525	3	85.00% F1
[43]	LSTM	COVID-19	Twitter, Facebook, Youtube	747,018	6	70.00% Acc
[42]	LSTM	COVID-19	Twitter	5.5 M	6	83.00% F1
[39]	ML Classifiers based on LSTM deep features	COVID-19	Twitter	685	2	94.01% Acc
[36]	BERT-based Models	COVID-19	Twitter	1.8 M	4	97.36% Acc
[40]	ML Classifiers	COVID-19	Twitter	8786	2	87.80% Acc
[41]	AraBERT-based Model	General	Twitter	779	2	87.70% Acc
[44]	BERT-based Models	COVID-19, Mental Health	Twitter	10,000	11	72.50% F1
[46]	ML Classifiers	Depression	Twitter	4542	3	82.39% Acc
[45]	ML Classifiers	Depression	Twitter	2722	2	87.50% Acc
[47]	LSTM	Depression	Facebook	10,000	>3	85.00% Acc
[48]	ML Classifiers with SMOTE	Diabetes	YouTube	4111	2	95.00% Acc
[50]	ML Classifiers	Cancer	Twitter	208	2	83.50% Acc
[49]	ML Classifiers	Influenza	Twitter	6300	2	89.06% Acc

3. Methodology and Proposed Approach

As previously stated, the aim of this study is to propose a framework that can be integrated as part of a disease surveillance system to help in detecting, tracking, and monitoring ILIs. This section describes our proposed system architecture for detecting ILI in people based on their Facebook postings using deep learning and NLP. The model's overall architecture is depicted in Figure 1. It consists of five modules designed to process and analyze Facebook post data. The initial module consists of data collection and annotation. The second module includes all preprocessing techniques used to work with the Arabic Algerian dialect. The third module encompasses FE techniques that turn text posts into meaningful representations. The fourth module utilizes oversampling and undersampling approaches to balance the dataset. Finally, the last module is related to the suggested deep learning model for sentiment classification. The subsequent subsections provide a full description of each module.

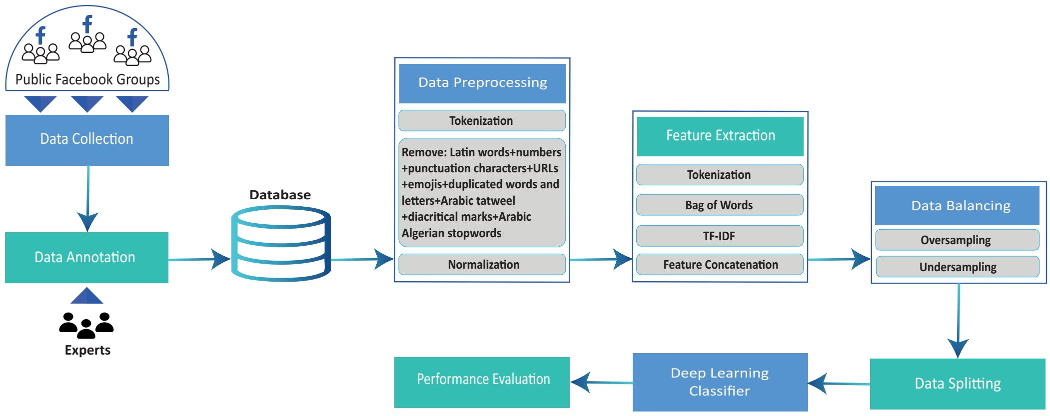


Figure 1. Adopted Methodology.

3.1. Data Collection

The data were collected from the most popular public Facebook groups in Algeria concerned with diseases and health issues. In each group, individuals express their health concerns (through wall posts) in order to receive medical advice or treatment from medical professionals or even non-medical group members. One of the benefits of using Facebook groups as a data source is that they provide data for a specific region in a specific language and area of interest, which facilitates data collection.

During the collection process, only textual content was retained; postings including photos or videos, as well as posts from group administrators, were discarded. Using multiple Facebook profiles, we collected data from March 2021 to 31 July 2021, until we obtained 21,885 postings.

The collected data consist of posts dating back to the inception of these Facebook groups on 24 April 2016. Since our analysis focuses on the detection of ILI, we have only included the data associated with the spread of COVID-19 in Algeria, i.e., from 01/01/2020 [51]. (see Figure 2).

On the other hand, it should be noted that the collected data respect the privacy and anonymity of each Facebook group's members and do not reveal the names of the posts' authors.

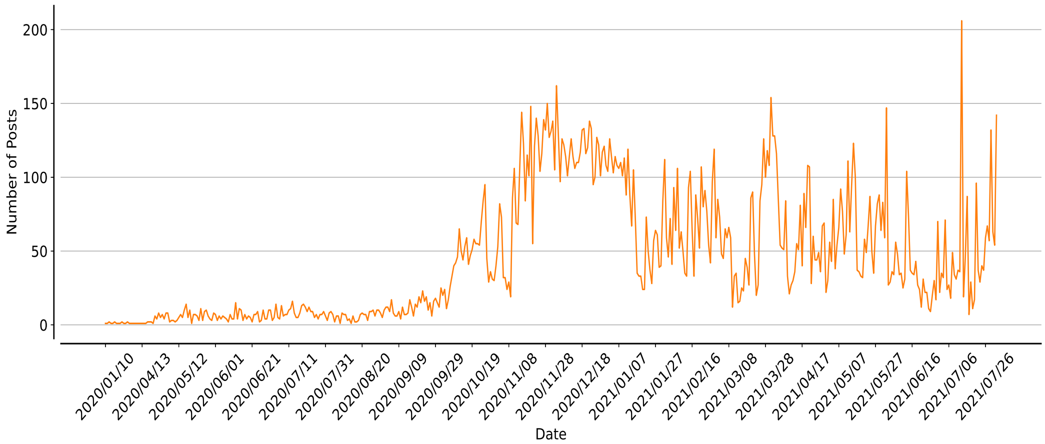


Figure 2. The volume of data collected from 01/01/2020 to 31/07/2021.

3.2. Data Annotation

After collecting data, the labeling process is performed based on the sentiment expressed in each Facebook post. This annotation stage is essential for preparing the data for the classification phase [52].

In our study, we manually annotated Facebook postings depending on the symptoms of ILI contained inside each post. They were categorized, with the aid of two annotators who are conversant with Algerian dialect, into the following three emotional categories:

- Positive:This category contains the postings whose authors claim they are experiencing ILI symptoms (such as fever, cough, sore throat, runny or stuffy nose, headaches, muscle aches, etc.) or new symptoms connected with COVID-19 (e.g., loss of taste or smell, difficulty breathing, chest pain).
- Negative-related: This category covers posts that do not indicate that the person is ill, but do provide medical advice or information regarding ILI symptoms.
- Unrelated: This category contains posts that are not related to ILI.

Table 2 illustrates examples of each of the above categories.

Table 2. Examples of posts in each sentiment class.

Class	Post in Arabic (Algerian Dialect)	Translated Post to English
Unrelated	نحتاج طبيب جلد مليح لنزع الشعر بالليزر تكون نتيجة مليحة شكون يعرف ولا تعرف	I need a good dermatologist for laser hair removal, with a good result, who knows a good doctor.
Negative-related	الكحة هى واحدة من الأعراض المُصاحبة لمرضٍ ما كالإنفلونَزا والرشح وغيرها من الأمراض المُنتشرة بالأخص فى فصل الشتاء وقد تكون علامةً وإشارة للشخص لينتبه لوجود أمرٍ خطير فى جسده	Cough is one of the symptoms that accompanies a disease such as influenza, cold and other diseases that are prevalent, especially in the winter season, and it may be a sign and signal for a person to be aware of the presence of something dangerous in his body.
Positive	السلام عليكم عندى السعال نسعل بزاف عندها يومين كاش دوا تع السعلة الله يجازيكم	Peace be upon you. I have a cough and I have been coughing a lot for two days. Is there a medicine for the cough, thank you.

The annotation procedure lasted around two months and yielded the following distribution of classes: Unrelated classes = 20,711 (94.63%), Positive classes = 936 (4.28%), and Negative-related classes = 238 (1.09%).

3.3. Data Analysis and Motivation

The collected data contain a wealth of information that can be used to benefit public health. Many diseases that are prevalent in Algerian society are mentioned in this information. According to N-gram analysis, the most common diseases and symptoms are: blood pressure (ضغط الدم), thyroid (الغدة الدرقية), nervous colon (قولون عصبى), shortness of breath (ضيق تنفس), blood sugar (سكر دم), and others.

In the context of our study, we compared the positive ILI cases in our database (Positive instances) with the COVID-19 cases recorded in Algeria by Johns Hopkins University's Center for Systems Science and Engineering (CSSE) [53]. We previously mentioned that COVID-19 has symptoms that are very similar to ILI, and some studies have even classified COVID-19 as an ILI [54,55]. Figure 3 shows the data from both databases normalized to the 0–1 scale.

The comparison of the graphs reveals that these two curves share certain similarities. Due to the paucity of data obtained from June to October 2020, the normalized curves exhibit a gap between instances from June to October 2020. However, there is a strong correlation between the two curves for the majority of the remaining months. Thus, we may conclude that the positive ILI patients in our dataset were related to the two waves of COVID-19 in Algeria. The first wave of COVID-19 began in October 2020 and ended in March 2021, while the second wave began in May 2021 and peaked in late July of the same year. The aforementioned investigations inspire us to present a sentiment classification system that detects ILI cases and contributes to the field of public health through intelligent systems for disease surveillance.

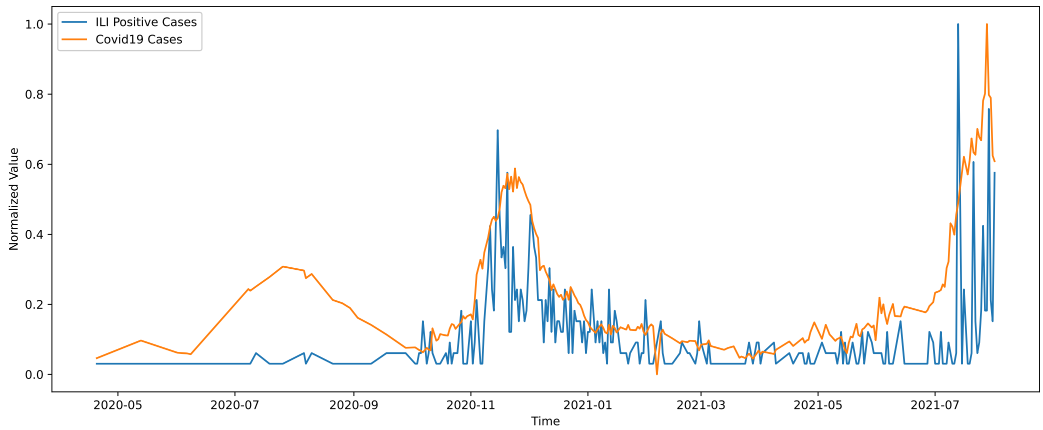


Figure 3. The positive ILI cases observed in our data compared to the COVID-19 cases from 20/04/2020 to 01/08/2021.

3.4. Data Preprocessing

Preprocessing is an important step in sentiment analysis [56]. At this stage, we eliminate all irrelevant and noisy data from the raw Facebook posts used in the sentiment classification process.

Each post was tokenized using N-grams (unigram) in order to facilitate the preparation of the raw data. Both character and word tokenization were considered. We will refer to them as Character-Tokenization and Word-Tokenization, respectively. In NLP, N-grams are sequences of N consecutive words (or characters) retrieved from textual data [57], where $N = 1$ corresponds to the use of uni-grams, $N = 2$ to bi-grams, $N = 3$ to tri-grams, etc.

Since Arabic is most commonly used to express opinions on Facebook in Algeria, all Latin letters and words were removed. In addition, we removed any numerals, punctuation, URLs, emojis, and repetitive words and letters from the same Facebook post. Additionally, any text posts with fewer than three words were deleted.

In addition, we eliminated Arabic stopwords (1574 words [58,59]) that do not contribute significantly to the meaning of the post. Furthermore, we suggest a new list of Arabic stopwords (400 words) based on the Algerian Arabic dialect that should likewise be eliminated.

We also converted some Arabic letters to another form (normalization). For example, ″لآ“،″لأ“،″لإ“ were converted to ″لا“, ″ى“ was converted to ″يـ“, and ″ه“ was converted to ″ة“. Moreover, for each word in the post, we removed Arabic tatweel (lengthening) and all Arabic diacritical marks (fatHah, kasrah, dhammah, shaddah, sukoon).

Before this phase, there were 21,885 raw data postings; after preprocessing, 1519 were eliminated, resulting in 20,366 posts.

The preceding preprocessing steps were applied to each Facebook post. Table 3 illustrates a data preprocessing application.

Table 3. Data preprocessing outcome on one Facebook post.

Before Data Preprocessing Phase
انا، عندى، فقدان، حساسة، الشم، والذوق، مع، انو، معنديش، حرارة، مرتفعة، هل، انا، مصاب؟ (I, have, loss, sense, smell, and taste, with, that, I don't have, high, temperature, is, I, injured?)
After Data Preprocessing Phase
فقدان، حساسة، شم، ذوق، حرارة، مرتفعة، مصاب (Loss, sensitivity, smell, taste, temperature, high, injured)

3.5. Feature Engineering

Typically, before using text data in deep learning-based NLP models, feature representations for each text instance in the dataset should be generated or extracted. All Facebook posts are integer-encoded at two levels in this regard: word-level and character-level. We used several techniques, including Tokenization, Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and Feature Concatenation, to convert raw texts into numerical values, as follows:

Tokenization: Each word and character is represented by a distinct integer. Following that, the integer vector representations at the word and character levels were padded with zeros to have the same lengths of $Nw = 447$ and $Nc = 2973$, which correspond to the number of words and characters in the longest text post, respectively.

BoW: This is a straightforward FE technique that counts the occurrences of each word/character in textual data to generate a numerical feature vector. BoW is widely used for topic modeling, NLP, text classification, and information retrieval due to its simplicity and effectiveness [60–62]. The size of the resulting feature vector is determined by the number of words or characters in the data.

TF-IDF: The weight of each term (word, letter) in the document is calculated using TF-IDF to determine its importance and rarity [63]. This weight is given based on its term frequency (TF) and inverse document frequency (IDF), as described in the formula below:

$$\text{TF-IDF}(t, d) = \text{TF}_{t,d} \times \text{IDF}_{t,d,N} = frq_{t,d} \times \log\left(\frac{N}{Nd_t}\right) \tag{1}$$

where $frq_{t,d}$ is the frequency of term t in document d. N is the total number of documents in the corpus. Nd_t is the number of documents containing the term t.

Feature concatenation: In addition to the previously mentioned FE approaches, we suggest feature concatenation using three different combination schemes: (1) word tokenization and character tokenization; (2) tokenization and BoW features; (3) tokenization and TF-IDF features. All of these concatenation schemes operate on word representations and/or character representations under various N-grams, including uni-grams, bi-grams, and tri-grams.

The encoding representation does not capture syntactic and semantic word relationships within text sequences [64,65]. In order to learn a mapping between words/characters during training, a word and/or character embedding layer is employed to receive the feature vector. Based on a vocabulary size of 22,752 words (135 characters), the embedding layer will generate a dense vector with dimensions $S \times E$, where S denotes the size of the feature vector and E represents the output embedding dimension. Thus, words and/or characters with similar meanings and common contexts will be mapped closely together in the vector space.

3.6. Data Balancing

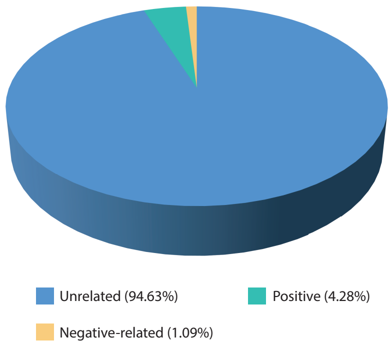
Significant patient information and medical history are stored in healthcare databases. The statistics reveal that the number of diagnosed disease cases (positive) has always been less than the number of healthy instances (negative) [66]. Interestingly, this also holds true for our obtained data, where the number of people suspected of having influenza is significantly smaller than the number of healthy people (see Figure 4a). This indicates that the data set collected is imbalanced in terms of class distribution. Using imbalanced data to train sentiment classification models, according to numerous studies [67,68], may result in erroneous precision and biased predictions.

Re-sampling methods, including undersampling and oversampling techniques, are among the most effective strategies that have been widely used in the literature to address the problem of imbalanced data [69–71]. Simply put, undersampling methods remove samples from the majority class, whereas oversampling methods increase the number of samples in the minority class [70].

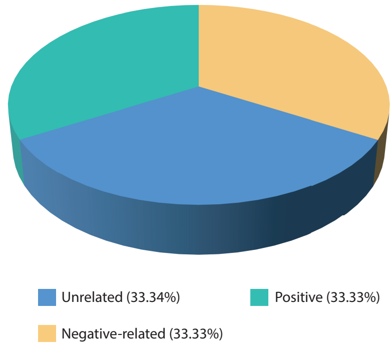
In our case, SMOTE [72,73] and Random Over Sampling (ROS) are used to oversample the "Positive" and "Negative-related" classes. On the other hand, from the Unrelated class that represents the majority class, we randomly selected 3000 instances to make the size of the three classes equal, as can be seen in Table 4.

The primary distinction between the two oversampling methods is that ROS is the most basic oversampling technique, in which minority class samples are randomly replicated. SMOTE, on the other hand, generates synthetic instances of the minority class along the line connecting this minority class to its nearest neighbor [72].

Figure 4 depicts the ratio of sentiment classes before and after applying balancing methods.



(**a**) Sentiment classes without balancing



(**b**) Sentiment classes with balancing

Figure 4. Ratio of sentiment classes with and without balancing.

Table 4. The number of instances for each class after SMOTE, ROS oversampling.

	Positive	Negative-Related	Unrelated	Total
Imbalanced	927	238	3000	4165
SMOTE	3000	3000	3000	9000
ROS	3000	3000	3000	9000

3.7. Sentiment Classification Using Deep Learning Model

In this study, we propose a deep learning model based on a convolutional neural network (CNN) [74] for ILI detection in Algerian Facebook posts.

Although CNNs have primarily been used in computer vision, they have also been used in NLP and produced impressive results [75–77]. CNNs can capture advanced features and handle input data in multiple dimensions. 2D-CNN and 3D-CNN are the most commonly used computer vision algorithms for images and video. Concurrently, 1D-CNN is used for 1D-signal processing, including biomedical data classification, speech recognition, structural health monitoring, and so on [78], as well as NLP [75,79].

Figure 5 depicts a graphical representation of the proposed 1D-CNN-based deep learning model. The proposed model's 13 layers include an input layer, an embedding layer, three 1D-convolutional layers, two max-pooling layers, four dropout layers, a global max-pooling layer, and a fully connected layer.

The input layer of our CNN model accepts each post as an integer-encoded vector. The embedding layer obtains the integer vector representation of S dimensions in order to map each word/character of a text post to an E-dimensional feature vector, producing a $S \times E$ matrix, where E represents the embedding dimension.

The output embedding matrix $S \times E$, followed by a dropout layer of 0.2, is then fed to the first 1D-convolutional layer with a filter size of 128 and a kernel size of 3. Faster than 2D-CNN [78], the kernel function in the 1D-CNN layer convolves the $S \times E$ matrix to extract hidden features and to detect local associations between adjacent characters. To capture the most relevant features and thus reduce the dimension of the preceding layer, the features from the first 1D-CNN layer are transmitted to the 1D max pooling layer, which is then followed by a dropout layer with a dropout rate of 0.2 to prevent overfitting.

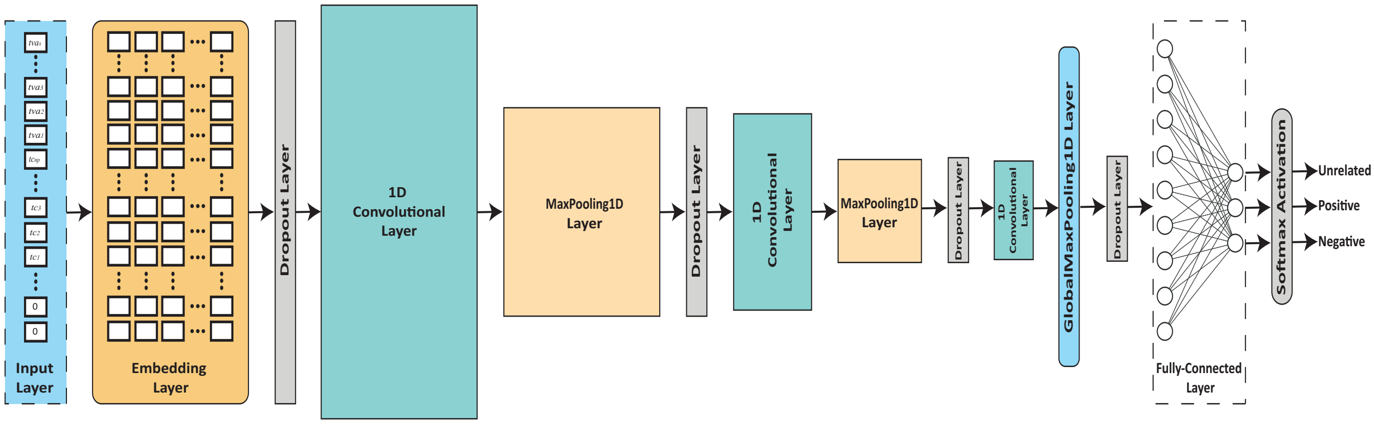


Figure 5. Architecture of the proposed 1D-CNN-based deep learning model.

To extract deeper features, an additional sequence of layers consisting of a 1D-convolutional layer with 64 filters, a 1D max pooling layer, and a dropout layer is added to the proposed model. The output of these layers is then transferred to the third 1D-CNN layer, which has a filter size of 16 and is followed by a global max pooling and dropout layer in order to reduce the network's complexity.

The final layer is the dense layer, a fully connected layer with the softmax activation function. As there are three classes, Positive, Negative-related, and Unrelated, the softmax function evaluates the probability value to return the class with the largest value.

It is important to note that several empirical attempts were made before settling on the 1D-CNN-based model, as evidenced by the results and explained in Section 4.

4. Experiments and Analysis

The aim of the conducted experiments is to evaluate the performance of the proposed 1D CNN-based model with various FE approaches and data balancing strategies. Moreover, the proposed model is compared to other baseline and state-of-the-art methods to evaluate its efficacy for sentiment classification.

A 5-fold cross-validation technique was used in our experimental study where each fold used for testing represents 20% of the data set, and the remaining 80% are used as training samples. Performance measures, including accuracy, precision, recall, F1-score, Receiver Operating Characteristics (ROC) curve, and Area Under the ROC Curve (AUC), were used to compare and evaluate the performances of the proposed model.

Furthermore, we took into account the embedding dimension (E), batch size, dropout rate, optimizer, and early stopping patience when tuning hyperparameters. Table 5 depicts the optimal parameter settings of our model.

We conducted all the experiments in Google Colab Pro (https://colab.research.google.com, accessed on 1 March 2021) Python 3 (CPU: Intel(R) Xeon(R) CPU @ 2.20 GHz; RAM: 25.46 GBs; Disk space: 166.83 GBs; GPU: Tesla P100-PCIE-16GB).

Table 5. Hyperparameter Setting.

Hyperparameter	Values Range	Optimal Value
Embedding dimension (E)	10, 20, 32, 64, 128	20
Batch size	32, 50, 64,128	128
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5	0.2
Optimizer	'SGD', 'RMSprop', 'adam', 'Nadam'	'adam'
Early stopping patience	1, 5, 10, 15, 20, 30	20

4.1. Evaluation Metrics

In this work, we adopt four evaluation metrics, including accuracy, precision, recall, and F1-score to evaluate the model's performance [80]. Each of these metrics is reported as an average of five folds. The value of each metric ranges between 0.0 (i.e., worst performance) and 1.0 (i.e., best performance), where the greater the value, the more efficient the model.

Accuracy is the proportion of correct predictions to total predictions. It is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

Precision refers to the proportion of positive predictions that actually belong to the positive class, which is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Recall denotes the proportion of real positives that are predicted correctly, calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

F1-score is defined as the harmonic mean of the precision and recall. It is considered as an essential performance evaluation measure for imbalanced data. F1-score is defined as follows:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

where, TP, FP, TN, and FN, in the above equations, refer to the number of True Positive, False Positive, True Negative, and False Negative cases, respectively.

4.2. Performance Results and Analysis

To account for the specificities of the Algerian Arabic dialect, we need to identify the best method extracting features and making sentiment classification more accurate. To identify an appropriate FE method for our dataset, we performed several feature concatenation schemes through various combinations between several feature engineering techniques, as illustrated in Table 6. Additionally, this experiment was conducted on imbalanced and oversampled data using SMOTE and ROS techniques to investigate their impact on sentiment classification.

Table 6 reveals a significant performance boost for the model using the oversampled data with ROS at all levels. Regardless of the FE approach, the proposed model achieved excellent results, with an average accuracy of 96.60%, as well as 96.60% precision, 96.50% recall, and 96.60% in F1-score while using feature concatenation between character-tokenization and word-level BoW with N-grams = 2.

Table 6. Performance of the proposed model with different FE techniques on imbalanced and oversampled data.

	#FE	**FE Technique**	**Level**	**N-Grams**	**Performance Metrics**			
					Accuracy	**Precision**	**Recall**	**F1-Score**
Imbalanced Dataset	1	Tokenization	Character	1	0.878	0.881	0.875	0.882
	2	Tokenization	Word	1	0.811	0.811	0.810	0.817
	3	Tokenization	Character + Word	1	0.807	0.808	0.807	0.815
	4	Tokenization + BoW	Character	1	0.883	0.886	0.881	0.885
	5	Tokenization + BoW	Character	1-2	0.892	0.895	0.888	0.895
	6	Tokenization + BoW	Character	1-3	0.891	0.894	0.888	0.894
	7	Tokenization + TF-IDF	Character	1	0.889	0.892	0.887	0.893
	8	Tokenization + TF-IDF	Character	1-2	0.896	0.898	0.893	0.898

Table 6. *Cont.*

	#FE	FE Technique	Level	N-Grams	Performance Metrics			
					Accuracy	Precision	Recall	F1-Score
	9	Tokenization + TF-IDF	Character	1-3	0.894	0.897	0.891	0.898
	10	Tokenization + BoW	Word	1	0.807	0.808	0.807	0.812
	11	Tokenization + BoW	Word	1-2	0.815	0.815	0.815	0.822
	12	Tokenization + BoW	Word	1-3	0.829	0.829	0.828	0.836
	13	Tokenization + TF-IDF	Word	1	0.822	0.825	0.821	0.828
	14	Tokenization + TF-IDF	Word	1-2	0.827	0.827	0.827	0.833
	15	Tokenization + TF-IDF	Word	1-3	0.821	0.822	0.820	0.828
	16	Tokenization + BoW	Character + Word	1	0.899	0.902	0.895	0.901
	17	Tokenization + BoW	Character + Word	1-2	0.894	0.896	0.891	0.897
	18	Tokenization + BoW	Character + Word	1-3	0.888	0.894	0.886	0.892
	19	Tokenization + TF-IDF	Character + Word	1	0.883	0.887	0.879	0.886
	20	Tokenization + TF-IDF	Character + Word	1-2	0.887	0.892	0.884	0.890
	21	Tokenization + TF-IDF	Character + Word	1-3	0.897	0.900	0.891	0.899
SMOTE	1	Tokenization	Character	1	0.884	0.886	0.883	0.882
	2	Tokenization	Word	1	0.706	0.710	0.702	0.698
	3	Tokenization	Character + Word	1	0.746	0.749	0.742	0.739
	4	Tokenization + BoW	Character	1	0.888	0.890	0.886	0.890
	5	Tokenization + BoW	Character	1-2	0.893	0.895	0.891	0.891
	6	Tokenization + BoW	Character	1-3	0.891	0.893	0.889	0.893
	7	Tokenization + TF-IDF	Character	1	0.888	0.890	0.885	0.893
	8	Tokenization + TF-IDF	Character	1-2	0.894	0.896	0.892	0.899
	9	Tokenization + TF-IDF	Character	1-3	0.895	0.897	0.892	0.895
	10	Tokenization + BoW	Word	1	0.719	0.722	0.714	0.706
	11	Tokenization + BoW	Word	1-2	0.721	0.723	0.718	0.713
	12	Tokenization + BoW	Word	1-3	0.738	0.745	0.735	0.740
	13	Tokenization + TF-IDF	Word	1	0.709	0.712	0.705	0.703
	14	Tokenization + TF-IDF	Word	1-2	0.727	0.733	0.721	0.712
	15	Tokenization + TF-IDF	Word	1-3	0.727	0.731	0.720	0.721
	16	Tokenization + BoW	Character + Word	1	0.893	0.894	0.890	0.893
	17	Tokenization + BoW	Character + Word	1-2	0.893	0.895	0.891	0.889
	18	Tokenization + BoW	Character + Word	1-3	0.891	0.893	0.888	0.892
	19	Tokenization + TF-IDF	Character + Word	1	0.898	0.900	0.896	0.899
	20	Tokenization + TF-IDF	Character + Word	1-2	0.894	0.897	0.893	0.890
	21	Tokenization + TF-IDF	Character + Word	1-3	0.893	0.894	0.890	0.891
ROS	1	Tokenization	Character	1	0.950	0.951	0.950	0.949
	2	Tokenization	Word	1	0.958	0.958	0.958	0.959
	3	Tokenization	Character + Word	1	0.950	0.951	0.950	0.952
	4	Tokenization + BoW	Character	1	0.960	0.960	0.960	0.959
	5	Tokenization + BoW	Character	1-2	0.963	0.963	0.963	0.965
	6	Tokenization + BoW	Character	1-3	0.958	0.959	0.958	0.961
	7	Tokenization + TF-IDF	Character	1	0.963	0.964	0.963	0.964
	8	Tokenization + TF-IDF	Character	1-2	0.964	0.964	0.964	0.966
	9	Tokenization + TF-IDF	Character	1-3	0.963	0.964	0.963	0.966
	10	Tokenization + BoW	Word	1	0.958	0.958	0.958	0.961
	11	Tokenization + BoW	Word	1-2	0.963	0.963	0.963	0.965
	12	Tokenization + BoW	Word	1-3	0.961	0.961	0.961	0.963
	13	Tokenization + TF-IDF	Word	1	0.961	0.962	0.961	0.962
	14	Tokenization + TF-IDF	Word	1-2	0.956	0.957	0.956	0.959
	15	Tokenization + TF-IDF	Word	1-3	0.964	0.964	0.963	0.966
	16	Tokenization + BoW	Character + Word	1	0.960	0.961	0.960	0.962
	17	Tokenization + BoW	Character + Word	1-2	**0.966**	**0.966**	**0.965**	**0.966**
	18	Tokenization + BoW	Character + Word	1-3	0.963	0.963	0.963	0.965
	19	Tokenization + TF-IDF	Character + Word	1	0.963	0.963	0.962	0.965
	20	Tokenization + TF-IDF	Character + Word	1-2	0.955	0.956	0.955	0.957
	21	Tokenization + TF-IDF	Character + Word	1-3	0.962	0.963	0.962	0.965

The values in bold are the best results for each metric.

Figure 6 displays the learning curves for the accuracy and loss of the proposed 1D-CNN-Based model during the training and validation phases while considering feature concatenation and data balance. These curves demonstrate that the proposed model was trained appropriately and that no overfitting was observed. For instance, the achieved training and validation accuracies were 98.50% and 96.70%, respectively, at epoch 155.

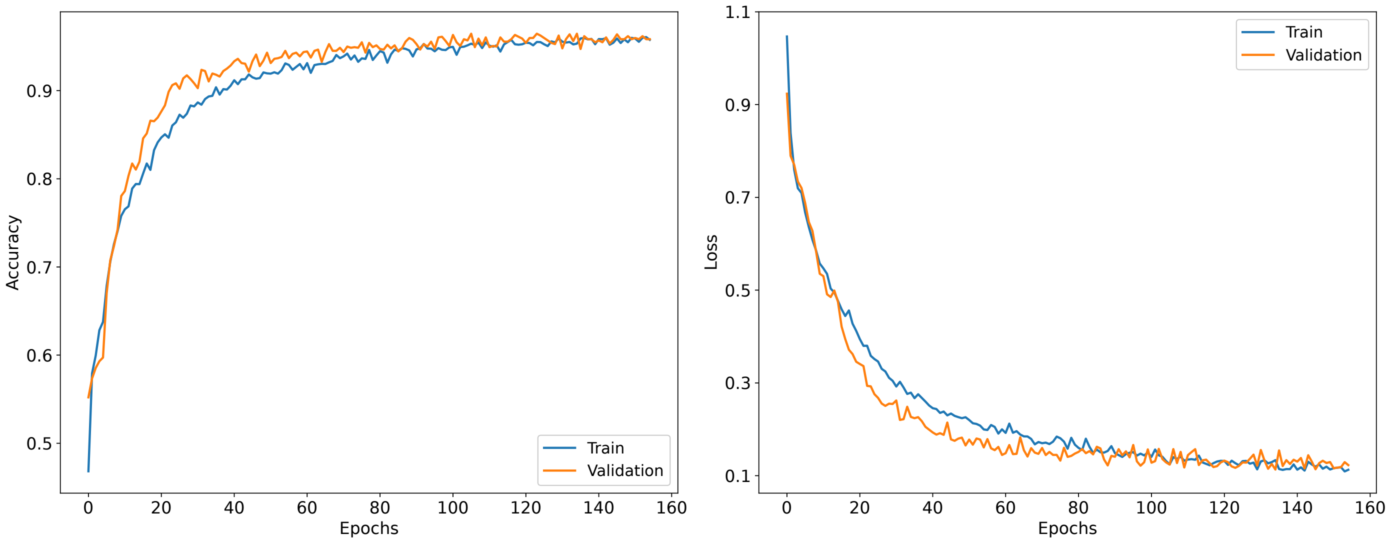


Figure 6. Accuracy and loss curves of our proposed 1D-CNN-based model.

On the other hand, a comparison of imbalanced and oversampled data using SMOTE reveals that neither has a significant advantage over the other. As we can see, the model trained on imbalanced data achieved an accuracy of 89.90%, precision of 90.20%, recall of 89.50%, and F1-score of 90.10%, while the model trained on SMOTE oversampled data achieved an accuracy of 89.80%, precision of 90.00%, recall of 89.60%, and 89.90% F1-score. In particular, the proposed model performed best with a feature concatenation that combines character-tokenization and word-level BoW using N-grams = 1 and imbalanced data. However, using SMOTE oversampled data, the proposed 1D-CNN-based model performs best with a feature concatenation combining character-tokenization and word-level TF-IDF using N-grams = 1.

Moreover, based on the results in Table 6, we graphically represented the proposed model's F1-score (suitable for imbalanced data), as illustrated in Figure 7. As such, it becomes clear that the performance of the proposed model on unbalanced and SMOTE oversampled data is negatively impacted when the FE process is based solely on words (see FE techniques: 2, 3, 10, 11, 12, 13, 14, 15). On ROS oversampled data, however, model results are unaffected when words and characters are utilized independently in the FE process. However, the performance improves when feature concatenation is employed.

As another way to evaluate these results, we present the confusion matrices depicted in Figure 8 corresponding to the proposed 1D-CNN-based model on the different datasets. A confusion matrix compares the true classes and the classes predicted by the proposed model. As shown in Figure 8a, the model using imbalanced data underperforms in identifying Positive and Negative-related sentiments, while it can achieve 92% correct predictions for the Unrelated class due to the availability of data in this category. Therefore, overfitting is most likely to occur in this case. The confusion matrix depicted in Figure 8b shows the results of the model with balanced data using SMOTE. As can be seen, 15% of positive instances were classified as Unrelated cases, which shows a misclassification issue which can be explained by the over-generalization problem related to SMOTE-based techniques.

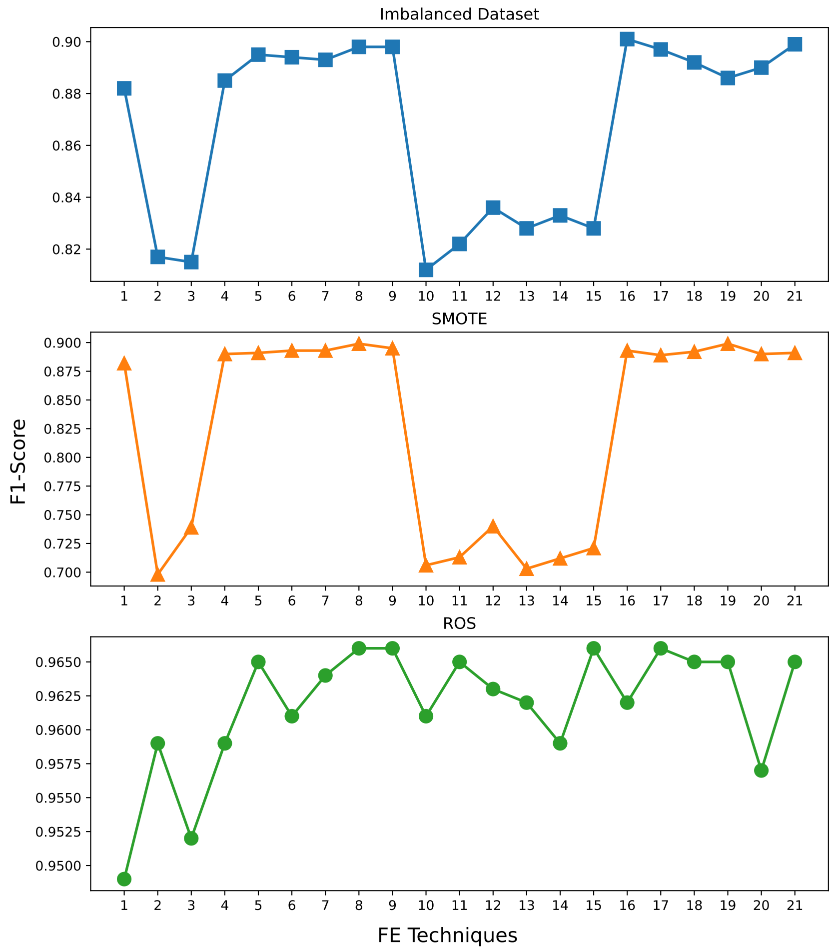
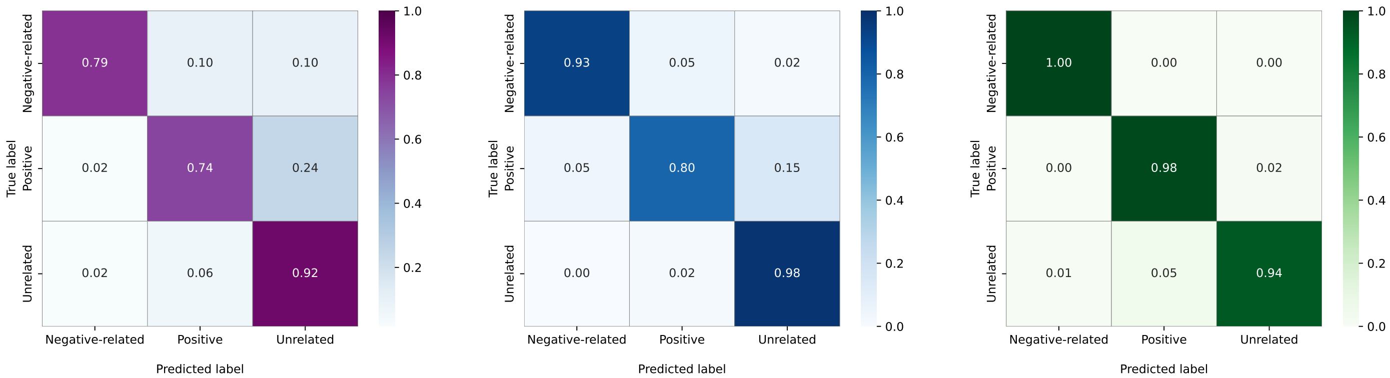


Figure 7. F1-score comparison with different FE techniques on imbalanced and oversampled data.

Finally, the confusion matrix for the model on ROS oversampled data (see Figure 8c) displays better and more accurate results in identifying all classes, resulting in high true positive rates ([94, 100%]) for Negative-related, Positive, and Unrelated sentiments.



(**a**) Imbalanced Dataset. (**b**) SMOTE. (**c**) ROS.

Figure 8. Confusion matrix of the proposed 1D-CNN-based model on imbalanced and oversampled data.

4.3. Comparison with Baselines

To show the validity and the effectiveness of the proposed 1D-CNN-based model, the performance of our algorithm was compared with the following sentiment classification baselines:

- LSTM is a type of recurrent neural network that uses different gates to learn long-term dependencies. It has been widely used for several sentiment classification tasks [81]. In this study, this model uses one LSTM layer with 128 neurons;
- GRU is a simpler and faster version of LSTM used widely in sequence problems. It consists of two gated functions: an update gate and a reset gate. The architecture of this model consists of one GRU layer with 64 neurons;
- BiLSTM is a sequence processing model with two LSTMs, one of which processes sequence data forward and the other backward. For this model, we use one bidirectional LSTM layer with 64 neurons;
- 1D-CNN is a feed-forward artificial neural network [82] that has been successfully used in various tasks related to NLP due to its remarkable ability to extract syntactic and semantic features. The architecture of this baseline consists of one 1D-CNN layer with 64 neurons, MaxPooling1D layer, and Flatten layer;
- 1D-CNN + LSTM is a hybrid deep learning model constructed by CNN and LSTM networks and thus combines the advantages of these two networks. In this model, we use the same layers in a 1D-CNN baseline with 128 neurons, followed by an LSTM layer with 64 neurons.

In order to obtain unbiased outcomes, in each baseline model, the data are oversampled using ROS, and the same feature concatenation is adopted, combining character tokenization and word-level BoW using N-grams = 2. Furthermore, all the above models incorporate an embedding layer and one dropout layer before the fully-connected dense layer with a softmax activation function, as described before in Figure 5. Additionally, we train each baseline using the same hyperparameters setting (see Table 5).

Table 7 compares the performance results of the proposed 1D-CNN-based model to those of the five baseline models. As can be seen, the proposed 1D-CNN-based model outperforms all the previously mentioned baseline models across all evaluation metrics. Particularly, according to accuracy, our model outperforms LSTM by 16.50%, GRU by 18.80%, BiLSTM by 34.90%,1D-CNN by 1.30%, and 1D-CNN+LSTM by 17.70%. Furthermore, the results show the effectiveness of all CNN-based models, including the 1D-CNN baseline model, compared to other methods and confirm the superior ability of CNN models in extracting the most discriminative features.

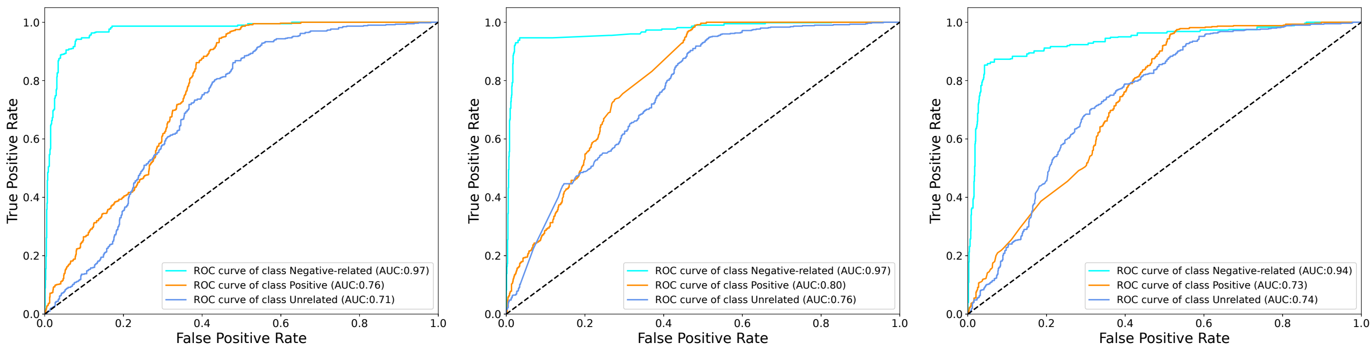
Table 7. Performance comparison of the proposed model with different baselines.

Model	Performance Metrics			
	Accuracy	**Precision**	**Recall**	**F1-Score**
LSTM	0.801	0.858	0.741	0.761
GRU	0.778	0.870	0.666	0.685
BiLSTM	0.617	0.840	0.464	0.490
1D-CNN	0.953	0.953	0.953	0.955
1D-CNN+LSTM	0.789	0.866	0.685	0.713
Proposed 1D-CNN-based model	**0.966**	**0.966**	**0.965**	**0.966**

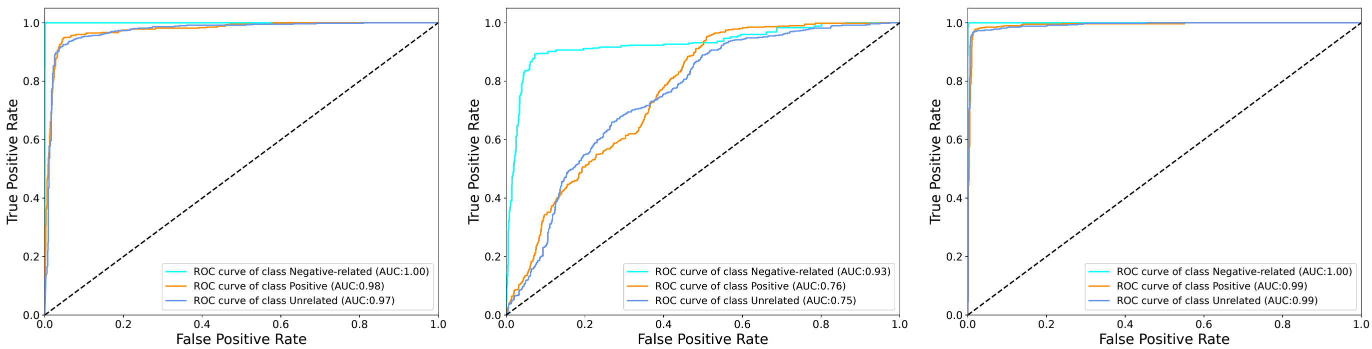
The values in bold are the best results for each metric.

With an accuracy of 61.70%, the BiLSTM baseline demonstrates the futility of using backward features. LSTM and GRU outperformed BiLSTM, with LSTM achieving the best results with an accuracy of 80.10%. The above results motivated us to combine LSTM and 1D-CNN (1D-CNN + LSTM) to improve sentiment classification performance. However, the obtained results did not show the expected improvement. Therefore, we focused our research on the 1D-CNN model by introducing more 1D-convolution layers, which resulted in the proposed 1D-CNN-based architecture in Figure 5.

To further evaluate the models, we generated the corresponding ROC curves to graphically represent and compare their performance (see Figure 9). The mean AUC was calculated as 0.81, 0.84, 0.80, 0.98, 0.81, and 0.99 for the LSTM, GRU, BiLSTM, 1D-CNN, 1D-CNN+LSTM, and the proposed model, respectively. This clearly demonstrates that the proposed 1D-CNN-based model outperforms the other baseline methods.



(**a**) LSTM (**b**) GRU (**c**) BiLSTM



(**d**) 1D-CNN (**e**) 1D-CNN + LSTM (**f**) Proposed 1D-CNN-based model

Figure 9. ROC curve comparison for (**a**) LSTM, (**b**) GRU, (**c**) BiLSTM, (**d**) 1D-CNN, (**e**) 1D-CNN+LSTM, and (**f**) the proposed 1D-CNN-based model.

4.4. Comparison with the State-of-the-Art Models

As shown in Table 8, we compared our proposed model with other state-of-the-art methods, including LSTM [42], SVM Bigram-TF-IDF [35], SVM Trigram-TF-IDF [48], Naive Bayes (NB) [49], and Random Forest (RF) [50]. The comparison was conducted using both unbalanced and balanced data based on ROS. As can be observed, the majority of comparison methods are based on traditional ML. This can be explained by the good performance of these algorithms in several sentiment classification studies, as shown in the section describing related work (see Table 1). These works are predominately based on simple feature extraction techniques, such as TF-IDF and tokenization, using various N-grams.

When balanced data are used, the results in Table 8 show that SVM-based methods perform very well, demonstrating their effectiveness in approaching sentiment classification problems [83,84], particularly in SVM Bigram-TF-IDF [35]. With an accuracy of 96.70%, the latter performed very similarly to our proposed 1D-CNN-based model, which had an accuracy of 96.60%. Other models, such as LSTM [42], SVM Trigram-TF-IDF [48], and RF [50], also performed well and were very close to each other. In contrast, the NB model has significant shortcomings when it comes to resolving the classification problem. Figure 10 depicts the confusion matrix of each model on oversampled data, which shows more details on the classification abilities of each method.

On the other hand, model comparison on imbalanced data revealed a clear difference between our proposed model and the other state-of-the-art methods, with our proposed model outperforming LSTM [42] by 7.10%, SVM Bigram-TF-IDF [35] by 9.20%, SVM Trigram-TF-IDF [48] by 14.10%, NB [49] by 62.40%, and RF [50] by 15.60%. Therefore, these results show the superiority of our proposed 1D-CNN model over the other models.

Table 8. Performance comparison of the proposed model with state-of-the-art methods.

	Model	**Performance Metrics**			
		Accuracy	**Precision**	**Recall**	**F1-Score**
ROS	LSTM [42]	0.948	0.949	0.948	0.951
	SVM Bigram-TF-IDF [35]	**0.967**	**0.968**	**0.967**	**0.967**
	SVM Trigram-TF-IDF [48]	0.955	0.960	0.955	0.955
	NB [49]	0.582	0.516	0.582	0.498
	RF [50]	0.937	0.947	0.937	0.937
	Proposed 1D-CNN-based model	0.966	0.966	0.965	0.966
Imbalanced Dataset	LSTM [42]	0.823	0.825	0.821	0.828
	SVM Bigram-TF-IDF [35]	0.802	0.819	0.802	0.766
	SVM Trigram-TF-IDF [48]	0.753	0.797	0.753	0.673
	NB [49]	0.270	0.526	0.270	0.143
	RF [50]	0.738	0.799	0.738	0.639
	Proposed 1D-CNN-based model	**0.894**	**0.896**	**0.891**	**0.897**

The values in bold are the best results for each metric.

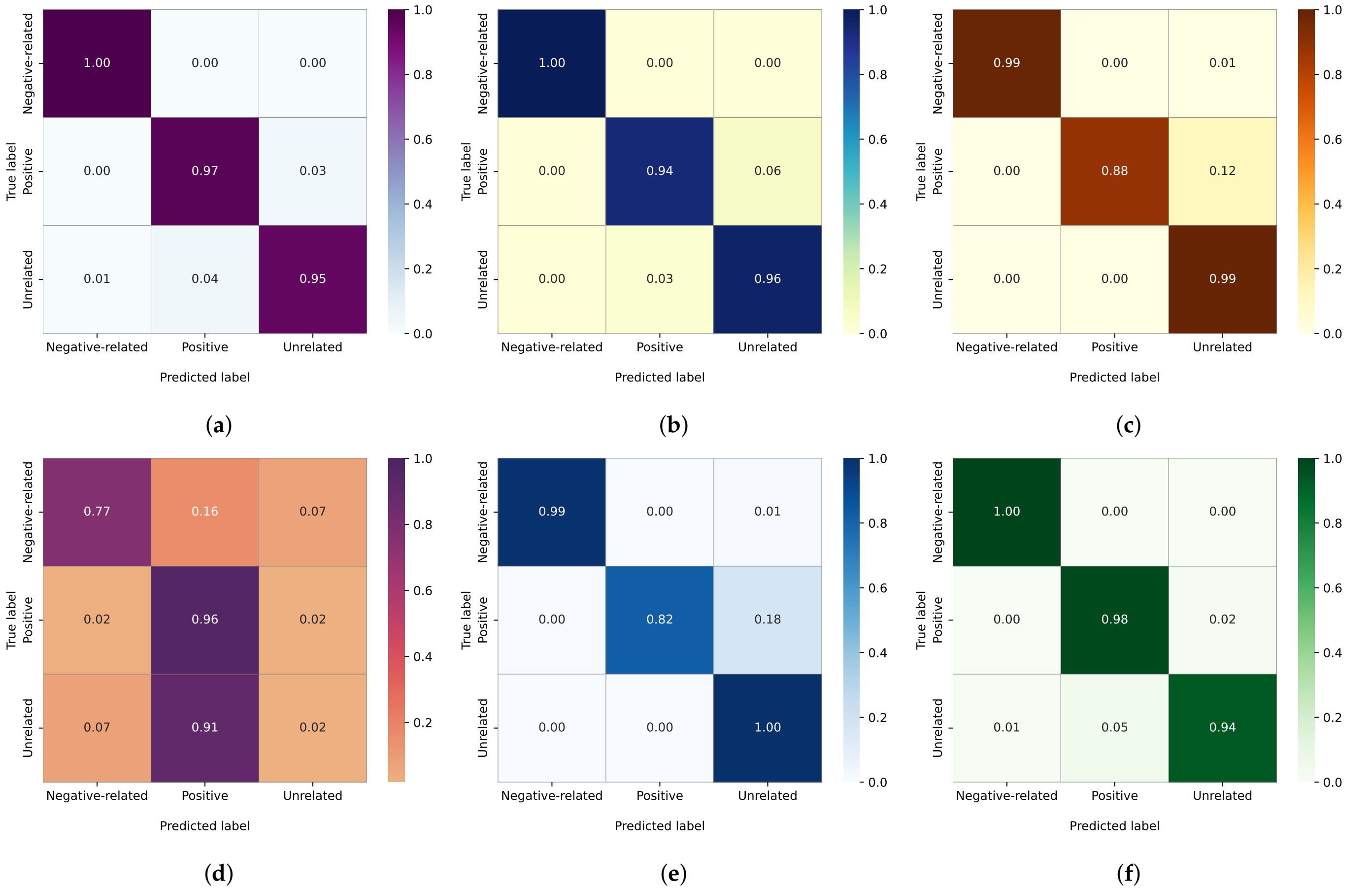


Figure 10. Confusion matrix of the proposed model compared to other state-of-the-art methods on oversampled data. (**a**) LSTM [42]. (**b**) SVM Bigram-TF-IDF [35]. (**c**) SVM Trigram-TF-IDF [48]. (**d**) NB[49]. (**e**) RF [50]. (**f**) Proposed 1D-CNN-based model.

5. Conclusions and Future Work

A framework for developing intelligent tools for disease surveillance based on social media posts is described in this paper. Core components of the proposed framework are the generation of a large dataset or corpus from Facebook posts written in the Algerian Arabic dialect and a multi-classification model based on 1D-CNN and sentiment analysis. Advanced NLP techniques were used to accurately analyze sentiments during an intensive data collection, labelling, and preparation task that led to the creation of the dataset.

Furthermore, to extract features from text data, we suggested using feature concatenation schemes that combine widely-used feature engineering techniques. In addition, ROS and SMOTE oversampling techniques were used to address the data imbalance problem. After data preprocessing, the proposed 1D-CNN classification model is a 13-layer deep learning model that has been trained and tested on the generated corpus. The experimental results demonstrate the effectiveness of the methods used for feature extraction and data balancing, and the proposed model achieved high performance with an average accuracy of 96.60% compared with the most popular models used in similar contexts such as SVM, BiLSTM, LSTM, and GRU. We intend to expand the current study to include the detection of even more diseases, which will benefit public health systems, as part of our future work. In addition, we plan to include other Arabic dialects in the proposed classification system. Combining our proposed model with a real-time data collection system to produce an online monitoring system would also be an interesting attempt.

Author Contributions: Conceptualization, A.B.; methodology, A.B. and S.M.; software, A.B.; validation, S.M. and K.A.; investigation, A.B. and K.A.; data curation, A.B. and K.A.; writing—original draft preparation, A.B.; writing—review and editing, S.M.; visualization, A.B.; supervision, A.B. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R196), Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All of the data and program code used in this study are available at the public repository: https://github.com/boulesnane/ILI-Detection (accessed on 1 November 2022).

Acknowledgments: The authors would like to acknowledge the Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R196), Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.

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

References

1. Rathore, A.K.; Kar, A.K.; Ilavarasan, P.V. Social Media Analytics: Literature Review and Directions for Future Research. *Decis. Anal.* **2017**, *14*, 229–249. [CrossRef]
2. Alalwan, A.A.; Rana, N.P.; Dwivedi, Y.K.; Algharabat, R. Social media in marketing: A review and analysis of the existing literature. *Telemat. Inform.* **2017**, *34*, 1177–1190. [CrossRef]
3. Anstead, N.; O'Loughlin, B. Social Media Analysis and Public Opinion: The 2010 UK General Election. *J. Comput.-Mediat. Commun.* **2014**, *20*, 204–220. [CrossRef]
4. Zeng, B.; Gerritsen, R. What do we know about social media in tourism? A review. *Tour. Manag. Perspect.* **2014**, *10*, 27–36. [CrossRef]
5. Yang, F.C.; Lee, A.J.; Kuo, S.C. Mining Health Social Media with Sentiment Analysis. *J. Med. Syst.* **2016**, *40*, 236. [CrossRef]
6. Haber, I.E.; Toth, M.; Hajdu, R.; Haber, K.; Pinter, G. Exploring Public Opinions on Renewable Energy by Using Conventional Methods and Social Media Analysis. *Energies* **2021**, *14*, 3089. [CrossRef]
7. Corbett, J.; Savarimuthu, B.T.R. From tweets to insights: A social media analysis of the emotion discourse of sustainable energy in the United States. *Energy Res. Soc. Sci.* **2022**, *89*, 102515. [CrossRef]
8. DataReportal. Digital 2022: Global Overview Report. 2022. Available online: https://datareportal.com/reports/digital-2022-global-overview-report (accessed on 1 September 2022).
9. DataReportal. Digital 2022: Algeria. 2022. Available online: https://datareportal.com/reports/digital-2022-algeria (accessed on 1 September 2022).
10. CDC. Overview of Influenza Surveillance in United States. USA: Department of Health and Human Services, Center for Disease Control. 2020. Available online: https://www.cdc.gov/flu/weekly/overview.htm (accessed on 8 February 2021).
11. Guan, W.-J.; Ni, Z.-Y.; Hu, Y.; Liang, W.-H.; Ou, C.-Q.; He, J.-X.; Liu, L.; Shan, H.; Lei, C.-L.; Hui, D.S.; et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [CrossRef]
12. Murtas, R.; Decarli, A.; Russo, A.G. Trend of pneumonia diagnosis in emergency departments as a COVID-19 surveillance system: A time series study. *BMJ Open* **2021**, *11*, e044388. [CrossRef]

13. Rustam, F.; Khalid, M.; Aslam, W.; Rupapara, V.; Mehmood, A.; Choi, G.S. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLOS ONE* **2021**, *16*, e0245909. [CrossRef]
14. Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassanien, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Appl. Soft Comput.* **2020**, *97*, 106754. [CrossRef] [PubMed]
15. Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P.W.; Kim, J. COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 1003–1015. [CrossRef] [PubMed]
16. Lim, S.; Tucker, C.S.; Kumara, S. An unsupervised machine learning model for discovering latent infectious diseases using social media data. *J. Biomed. Inform.* **2017**, *66*, 82–94. [CrossRef] [PubMed]
17. García-Díaz, J.A.; Apolinario-Arzube, Ó.; Medina-Moreira, J.; Luna-Aveiga, H.; Lagos-Ortiz, K.; Valencia-García, R. Sentiment Analysis on Tweets related to infectious diseases in South America. In Proceedings of the Euro American Conference on Telematics and Information Systems, Fortaleza, Brazil, 12–15 November 2018. [CrossRef]
18. Babu, N.V.; Kanaga, E.G.M. Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. *SN Comput. Sci.* **2021**, *3*, 74. [CrossRef]
19. Hassan, A.U.; Hussain, J.; Hussain, M.; Sadiq, M.; Lee, S. Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. In Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 18–20 October 2017. [CrossRef]
20. Joshi, M.L.; Kanoongo, N. Depression detection using emotional artificial intelligence and machine learning: A closer review. *Mater. Today Proc.* **2022**, *58*, 217–226. [CrossRef]
21. Hinduja, S.; Afrin, M.; Mistry, S.; Krishna, A. Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100113. [CrossRef]
22. Sumathy, B.; Kumar, A.; Sungeetha, D.; Hashmi, A.; Saxena, A.; Shukla, P.K.; Nuagah, S.J. Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System. *Comput. Intell. Neurosci.* **2022**, *2022*, 5906797. [CrossRef]
23. Jain, V.K.; Kumar, S. Effective surveillance and predictive mapping of mosquito-borne diseases using social media. *J. Comput. Sci.* **2018**, *25*, 406–415. [CrossRef]
24. Gabarron, E.; Dechsling, A.; Skafle, I.; Nordahl-Hansen, A. Discussions of Asperger Syndrome on Social Media: Content and Sentiment Analysis on Twitter. *JMIR Form. Res.* **2022**, *6*, e32752. [CrossRef]
25. Amin, S.; Uddin, M.I.; Hassan, S.; Khan, A.; Nasser, N.; Alharbi, A.; Alyami, H. Recurrent Neural Networks With TF-IDF Embedding Technique for Detection and Classification in Tweets of Dengue Disease. *IEEE Access* **2020**, *8*, 131522–131533. [CrossRef]
26. Yousefinaghani, S.; Dara, R.; Poljak, Z.; Bernardo, T.M.; Sharif, S. The Assessment of Twitter's Potential for Outbreak Detection: Avian Influenza Case Study. *Sci. Rep.* **2019**, *9*, 18147. [CrossRef] [PubMed]
27. Zhang, F.; Luo, J.; Li, C.; Wang, X.; Zhao, Z. Detecting and Analyzing Influenza Epidemics with Social Media in China. In *Advances in Knowledge Discovery and Data Mining*; Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 90–101. [CrossRef]
28. Alessa, A.; Faezipour, M. A review of influenza detection and prediction through social networking sites. *Theor. Biol. Med. Model.* **2018**, *15*. [CrossRef] [PubMed]
29. Jain, V.K.; Kumar, S. An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter. *Procedia Comput. Sci.* **2015**, *70*, 801–807. [CrossRef]
30. Zuccon, G.; Khanna, S.; Nguyen, A.; Boyle, J.; Hamlet, M.; Cameron, M. Automatic detection of tweets reporting cases of influenza like illnesses in Australia. *Health Inf. Sci. Syst.* **2015**, *3*, S4. [CrossRef] [PubMed]
31. Alkouz, B.; Aghbari, Z.A.; Al-Garadi, M.A.; Sarker, A. Deepluenza: Deep learning for influenza detection from Twitter. *Expert Syst. Appl.* **2022**, *198*, 116845. [CrossRef]
32. Asiri, E.; Khalifa, M.; Shabir, S.A.; Hossain, M.N.; Iqbal, U.; Househ, M. Sharing sensitive health information through social media in the Arab world. *Int. J. Qual. Health Care* **2016**, *29*, 68–74. [CrossRef]
33. Birjali, M.; Kasri, M.; Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl.-Based Syst.* **2021**, *226*, 107134. [CrossRef]
34. Binkheder, S.; Aldekhyyel, R.N.; AlMogbel, A.; Al-Twairesh, N.; Alhumaid, N.; Aldekhyyel, S.N.; Jamal, A.A. Public Perceptions around mHealth Applications during COVID-19 Pandemic: A Network and Sentiment Analysis of Tweets in Saudi Arabia. *Int. J. Environ. Res. Public Health* **2021**, *18*, 13388. [CrossRef]
35. Aljameel, S.S.; Alabbad, D.A.; Alzahrani, N.A.; Alqarni, S.M.; Alamoudi, F.A.; Babili, L.M.; Aljaafary, S.K.; Alshamrani, F.M. A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia. *Int. J. Environ. Res. Public Health* **2020**, *18*, 218. [CrossRef]
36. Essam, N.; Moussa, A.M.; Elsayed, K.M.; Abdou, S.; Rashwan, M.; Khatoon, S.; Hasan, M.M.; Asif, A.; Alshamari, M.A. Location Analysis for Arabic COVID-19 Twitter Data Using Enhanced Dialect Identification Models. *Appl. Sci.* **2021**, *11*, 11328. [CrossRef]
37. Addawood, A. Coronavirus: Public Arabic Twitter Data Set. 2020. Available online: https://openreview.net/forum?id=ZxjFAfD0pSy (accessed on 22 October 2022).

38. Zaidan, O.; Callison-Burch, C. The arabic online commentary dataset: An annotated dataset of informal arabic with high dialectal content. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 37–41.
39. Alabrah, A.; Alawadh, H.M.; Okon, O.D.; Meraj, T.; Rauf, H.T. Gulf Countries' Citizens' Acceptance of COVID-19 Vaccines—A Machine Learning Approach. *Mathematics* **2022**, *10*, 467. [CrossRef]
40. Alqurashi, S.; Hamoui, B.; Alashaikh, A.; Alhindi, A.; Alanazi, E. Eating Garlic Prevents COVID-19 Infection: Detecting Misinformation on the Arabic Content of Twitter. *arXiv* **2021**, arXiv:2101.05626.
41. Albalawi, Y.; Nikolov, N.S.; Buckley, J. Pretrained Transformer Language Models Versus Pretrained Word Embeddings for the Detection of Accurate Health Information on Arabic Social Media: Comparative Study. *JMIR Form. Res.* **2022**, *6*, e34834. [CrossRef] [PubMed]
42. Al-Laith, A.; Alenezi, M. Monitoring People's Emotions and Symptoms from Arabic Tweets during the COVID-19 Pandemic. *Information* **2021**, *12*, 86. [CrossRef]
43. Ghanem, A.; Asaad, C.; Hafidi, H.; Moukafih, Y.; Guermah, B.; Sbihi, N.; Zakroum, M.; Ghogho, M.; Dairi, M.; Cherqaoui, M.; et al. Real-Time Infoveillance of Moroccan Social Media Users' Sentiments towards the COVID-19 Pandemic and Its Management. *Int. J. Environ. Res. Public Health* **2021**, *18*, 12172. [CrossRef]
44. Alturayeif, N.; Luqman, H. Fine-Grained Sentiment Analysis of Arabic COVID-19 Tweets Using BERT-Based Transformers and Dynamically Weighted Loss Function. *Appl. Sci.* **2021**, *11*, 10694. [CrossRef]
45. Almouzini, S.; khemakhem, M.; Alageel, A. Detecting Arabic Depressed Users from Twitter Data. *Procedia Comput. Sci.* **2019**, *163*, 257–265. [CrossRef]
46. Musleh, D.A.; Alkhales, T.A.; Almakki, R.A.; Alnajim, S.E.; Almarshad, S.K.; Alhasaniah, R.S.; Aljameel, S.S.; Almuqhim, A.A. Twitter Arabic Sentiment Analysis to Detect Depression Using Machine Learning. *Comput. Mater. Contin.* **2022**, *71*, 3463–3477. [CrossRef]
47. ElDin, D.M.; Hamed, M.; Eldeen, N. SentiNeural: A Depression Clustering Technique for Egyptian Women Sentiments. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*. [CrossRef]
48. Yafooz, W.M.; Alsaeedi, A. Sentimental Analysis on Health-Related Information with Improving Model Performance using Machine Learning. *J. Comput. Sci.* **2021**, *17*, 112–122. [CrossRef]
49. Baker, Q.; Shatnawi, F.; Rawashdeh, S.; Al-Smadi, M.; Jararweh, Y. Detecting Epidemic Diseases Using Sentiment Analysis of Arabic Tweets. *JUCS J. Univers. Comput. Sci.* **2020**, *26*, 50–70. [CrossRef]
50. Saeed, F.; Yafooz, W.M.S.; Al-Sarem, M.; Abdullah, E. Detecting Health-Related Rumors on Twitter using Machine Learning Methods. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*. [CrossRef]
51. Lounis, M. Epdemiology of coronavirus disease 2020 (COVID-19) in Algeria. *New Microbes New Infect.* **2021**, *39*, 100822. [CrossRef] [PubMed]
52. Al-Twairesh, N.; Al-Khalifa, H.; Al-Salman, A.; Al-Ohali, Y. AraSenTi-Tweet: A Corpus for Arabic Sentiment Analysis of Saudi Tweets. *Procedia Comput. Sci.* **2017**, *117*, 63–72. [CrossRef]
53. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [CrossRef]
54. Amin, M.T.; Fatema, K.; Arefin, S.; Hussain, F.; Bhowmik, D.R.; Hossain, M.S. Obesity, a major risk factor for immunity and severe outcomes of COVID-19. *Biosci. Rep.* **2021**, *41*, BSR20210979. [CrossRef]
55. Kumar, R.; Arora, R.; Bansal, V.; Sahayasheela, V.J.; Buckchash, H.; Imran, J.; Narayanan, N.; Pandian, G.N.; Raman, B. Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers. *medRxiv* **2020**. [CrossRef]
56. Symeonidis, S.; Effrosynidis, D.; Arampatzis, A. A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Syst. Appl.* **2018**, *110*, 298–310. [CrossRef]
57. Sidorov, G.; Velasquez, F.; Stamatatos, E.; Gelbukh, A.; Chanona-Hernández, L. Syntactic N-grams as machine learning features for natural language processing. *Expert Syst. Appl.* **2014**, *41*, 853–860. [CrossRef]
58. El-Khair, I.A. Effects of stop words elimination for Arabic information retrieval: A comparative study. *Int. J. Comput. Inf. Sci.* **2006**, *4*, 119–133.
59. PyArabic. PyPI. Available online: https://pypi.org/project/PyArabic/ (accessed on 1 September 2021).
60. Qin, Z.; Cong, Y.; Wan, T. Topic modeling of Chinese language beyond a bag-of-words. *Comput. Speech Lang.* **2016**, *40*, 60–78. [CrossRef]
61. HaCohen-Kerner, Y.; Miller, D.; Yigal, Y. The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE* **2020**, *15*, e0232525. [CrossRef] [PubMed]
62. Passalis, N.; Tefas, A. Learning bag-of-embedded-words representations for textual information retrieval. *Pattern Recognit.* **2018**, *81*, 254–267. [CrossRef]
63. Zhang, W.; Yoshida, T.; Tang, X. A comparative study of TF* IDF, LSI and multi-words for text classification. *Expert Syst. Appl.* **2011**, *38*, 2758–2765. [CrossRef]
64. Lauriola, I.; Lavelli, A.; Aiolli, F. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing* **2022**, *470*, 443–456. [CrossRef]

65. Kumar, V.; Recupero, D.R.; Riboni, D.; Helaoui, R. Ensembling Classical Machine Learning and Deep Learning Approaches for Morbidity Identification From Clinical Notes. *IEEE Access* **2021**, *9*, 7107–7126. [CrossRef]
66. Kaur, H.; Pannu, H.S.; Malhi, A.K. A Systematic Review on Imbalanced Data Challenges in Machine Learning. *ACM Comput. Surv.* **2019**, *52*, 1–36. [CrossRef]
67. Singla, Z.; Randhawa, S.; Jain, S. Sentiment analysis of customer product reviews using machine learning. In Proceedings of the 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 23–24 June 2017; pp. 1–5. [CrossRef]
68. Tolba, M.; Ouadfel, S.; Meshoul, S. Hybrid ensemble approaches to online harassment detection in highly imbalanced data. *Expert Syst. Appl.* **2021**, *175*, 114751. [CrossRef]
69. Ramos-Pérez, I.; Arnaiz-González, Á.; Rodríguez, J.J.; García-Osorio, C. When is resampling beneficial for feature selection with imbalanced wide data? *Expert Syst. Appl.* **2022**, *188*, 116015. [CrossRef]
70. Liang, D.; Yi, B.; Cao, W.; Zheng, Q. Exploring ensemble oversampling method for imbalanced keyword extraction learning in policy text based on three-way decisions and SMOTE. *Expert Syst. Appl.* **2022**, *188*, 116051. [CrossRef]
71. Houssein, E.H.; Hassaballah, M.; Ibrahim, I.E.; AbdElminaam, D.S.; Wazery, Y.M. An automatic arrhythmia classification model based on improved Marine Predators Algorithm and Convolutions Neural Networks. *Expert Syst. Appl.* **2022**, *187*, 115936. [CrossRef]
72. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
73. Elreedy, D.; Atiya, A.F. A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Inf. Sci.* **2019**, *505*, 32–64. [CrossRef]
74. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–21. [CrossRef] [PubMed]
75. Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv* **2014**, arXiv:1408.5882.
76. Giménez, M.; Palanca, J.; Botti, V. Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis. *Neurocomputing* **2020**, *378*, 315–323. [CrossRef]
77. Conneau, A.; Schwenk, H.; Barrault, L.; Lecun, Y. Very Deep Convolutional Networks for Text Classification. *arXiv* **2016**, arXiv:1606.01781.
78. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [CrossRef]
79. Sharma, A.K.; Chaurasia, S.; Srivastava, D.K. Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec. *Procedia Comput. Sci.* **2020**, *167*, 1139–1147. [CrossRef]
80. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756.
81. Joseph, J.; Vineetha, S.; Sobhana, N. A survey on deep learning based sentiment analysis. *Mater. Today Proc.* **2022**, *58*, 456–460. [CrossRef]
82. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
83. Li, X.; Li, J.; Wu, Y. A Global Optimization Approach to Multi-Polarity Sentiment Analysis. *PLoS ONE* **2015**, *10*, e0124672. [CrossRef] [PubMed]
84. AlBadani, B.; Shi, R.; Dong, J. A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM. *Appl. Syst. Innov.* **2022**, *5*, 13. [CrossRef]