International Journal of Disaster Risk Reduction 96 (2023) 103972

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Multi-modal mining of crowd-sourced data: Efficient provision of humanitarian aid to remote regions affected by natural disasters

Sadegh Khanmohammadi [a], Emad Golafshani [a], Yu Bai [a], Heng Li [b], Milad Bazli [c], Mehrdad Arashpour [a,*]

[a] Department of Civil Engineering, Monash University, Melbourne, Australia
[b] Hong Kong Polytechnic University, Hong Kong, China
[c] College of Engineering, IT and Environment, Charles Darwin University, Australia

ARTICLE INFO

Keywords:
Humanitarian aid
Data mining
Artificial intelligence (AI)
Deep neural networks (DNN)
SHapley additive exPlanations (SHAP)
Generative adversarial networks (GAN)

ABSTRACT

Data mining applications have the potential to address current deficiencies in the provision of humanitarian aid in natural disasters. Simultaneous text and image analysis in crowd-sourced data can improve the quality of humanitarian aid information. Specifically, we select Bidirectional Encoder Representations from Transformers (BERT) and its descendant ALBERT as pre-trained deep networks for the text modality, while we choose ConvNeXt, RegNet, and Faster RCNN for the image modality. The developed framework demonstrates its application in classifying humanitarian aid through three key aspects. Firstly, it illustrates the effective performance of ConvNeXt and BERT in the classification of humanitarian aid. Secondly, it investigates the efficiency of generative adversarial networks (GAN) in generating synthetic images for imbalanced input datasets. This approach improves the accuracy, precision, recall, and F1-score of the framework when applied to unseen test data. Finally, the study highlights the potential use of SHapley Additive exPlanations (SHAP) for interpreting the behaviour of the developed framework, supporting the timely classification of humanitarian aid information from crowd-sourced data after natural disasters.

1. Introduction

Early provision of humanitarian aid after a natural disaster (i.e., the golden window) plays an important role in reducing the number of casualties and fatalities [1]. Crowd-sourced data can provide valuable information for efficient and timely humanitarian aid [2,3]. In the first hours after a disaster occurs, information about casualties and missing people, needs, and availabilities are often posted on social media [4]. However, it is not easy to analyse crowd-sourced data due to the complexity of the humanitarian aid data obtained from social media [5]. Considering the overload of post-disaster information, classifying informative feeds based on their content and transferring them to suitable emergency responders are essential tasks [6,7]. Accurate classifications of humanitarian aid information in social media feeds are critical for immediate actions and saving lives [8].

Machine learning (ML) methods have shown promising performance in analysing crowd-sourced feeds [9]. Among various machine learning methods used by previous studies for such a purpose, deep learning (DL) methods have been strongly recommended by recent studies due to their superior performance in comparison with traditional ML models [10,11]. Crowd-sourced feeds from social

* Corresponding author. Monash University, Department of Civil Engineering, Vic, 3800, Australia.
E-mail addresses: sadegh.khanmohammadi@monash.edu (S. Khanmohammadi), emad.golafshani@monash.edu (E. Golafshani), yu.bai@monash.edu (Y. Bai), heng.li@polyu.edu.hk (H. Li), milad.bazli@cdu.edu.au (M. Bazli), Mehrdad.arashpour@monash.edu (M. Arashpour).

https://doi.org/10.1016/j.ijdrr.2023.103972
Received 3 January 2023; Received in revised form 28 August 2023; Accepted 29 August 2023
Available online 29 August 2023

media mainly consist of images and text, and both text and image feeds are important in providing more insightful humanitarian aid information. Consequently, a multi-modal analysis to process and classify both text and image extractions is highly desirable [12]. In addition, social media feeds generally contain imbalanced sources of data [13], which poses a challenge to DL-based classifiers [14,15]. The performance of a DL-based classifier is densely related to data quality and patterns, and imbalanced datasets can reduce the accuracy of the classifiers [16,17]. Furthermore, the complexity of DL models introduces challenges to interpreting models and understanding their behaviour. Due to the significance of the humanitarian aid issue, using black-box models is not rational [18]. Therefore, the current study aims to develop a framework to perform an interpretable multi-modal analysis of social media and solve the imbalanced dataset problem. To achieve the research aim, answers to the following questions need to be sought.

RQ1. Which DL architectures can analyse crowd-sourced data related to natural disasters in remote regions?

Shallow Natural Language Processing (NLP) networks, including Word2Vec [12] and FastText [19] in combination with very Deep Convolutional Networks (VGGNet) [20], and deep RESidual NETworks (ResNet) [21] have been employed in multimodality studies. However, few studies have evaluated the performance of recent NLP and image-processing models. To provide an insightful response to this question, the classification performance of two recent NLP models and two up-to-date image processing models are evaluated. Bidirectional Encoder Representations from Transformers (BERT) [22] and one of its decedents i.e. A Lite BERT (ALBERT) [23], as two NLP models, and RegNetY320 [24] and ConvNeXt [25], as two image processing models, are employed in this study. By using these models and their combinations, we develop four multi-modal models and compare their performances using an unseen test dataset. The selection of the mentioned algorithms is related to their outstanding performance, especially for social media content based on past studies [11,26,27].

RQ2. What is a possible solution for solving the imbalance dataset issue in crowd-sourced data?

To address the issue of the imbalance dataset, generative adversarial networks (GAN) which have the capability of generating synthetic images are considered. This technique can generate new data by using the available data to enrich classes with lower data records. The usefulness of GAN for imbalanced social media datasets has been investigated in past studies [28,29]. However, the GAN impact on multimodality research has not been explored. In this study, the impact of using GAN synthetic images in the model training process is studied by evaluating the models against an unseen test dataset. The imbalanced input dataset of social media content for humanitarian aid studies is related to the fact that sharing images of affected people after disaster is restricted by rules.

RQ3. How can we interpret the behaviour of complex deep learning networks in analysing crowd-sourced data?

Providing humanitarian aid in a post-disaster situation can preserve lives. All ML models that are employed for this critical issue should be properly interpreted before applications [18]. According to the black-box nature of ML models, the relationship between inputs and outputs could be discovered using an interpretation tool [30,31]. The interpretation of ML models on social media feeds and the interpretation of DL networks for images have been limited in past studies. Shapley Additive explanation (SHAP) is employed to answer the third research question. SHAP helps humanitarian organisations interpret the multi-modal models and explain data patterns and behaviours.

The main contributions of this paper can be summarized into three parts. First, methodological contributions that are related to developing state-of-the-art and accurate models. Second, theoretical contributions are associated with proposing synthetic data records to solve the imbalance issue of humanitarian aid datasets. Third, practical contributions that are connected with the interpretation of black box models. The paper is structured as follows. Section 2 is devoted to the relevant humanitarian aid research studies on single modality (text or image) and multi-modal analysis of crowd-sourced data. The methodology and materials of this research study are outlined in Section 3. The analysis and results are provided in Section 4. Research discussions are presented in Section 5. Finally, the conclusions, research limitations, and opportunities for future studies are provided in Section 6.

2. Background

Social media posts have been widely used to provide rapid situational awareness in disasters due to their real-time and human-centred nature [32]. People could offer or request humanitarian aid through social media platforms [33]. Although social media feeds provide rich datasets for research, the feeds frequently contain irrelevant content [34]. ML approaches appear to be powerful methods for extracting humanitarian aspects of crowd-sourced data during emergencies to alleviate disaster-affected people [35]. Previous humanitarian aid research on analysing crowd-sourced data by ML can be divided into three categories of text, image, and both combined based on their inputs. The following subsections provide background information on each mentioned category. The subsections provide information about the necessity of ML applications for crowd-sourced data, the important characteristics of crowd-sourced data as input to data-driven modelling, and ML advancements for providing humanitarian aid after natural disasters.

2.1. Textual modality

NLP approaches to extract humanitarian aid information from crowd-sourced data can be designated as the first category. One of the main recent trends is to classify information by NLP systems, as can be seen in the literature map shown in Fig. 1 (a). Furthermore, Fig. 1 (a) demonstrates that online social networking can provide valuable inputs for NLP systems. It is worth noting that Fig. 1 is an un-directional network. The keyword networks of natural language processing (for text modality analysis), image processing (for image modality analysis), and multi-modal analysis (analysis of text and image simultaneously are depicted in Fig. 1 a, b, and c, respectively. Table 1 gives a few recent NLP studies on information classification in a post-disaster situation. A few past studies have devel-

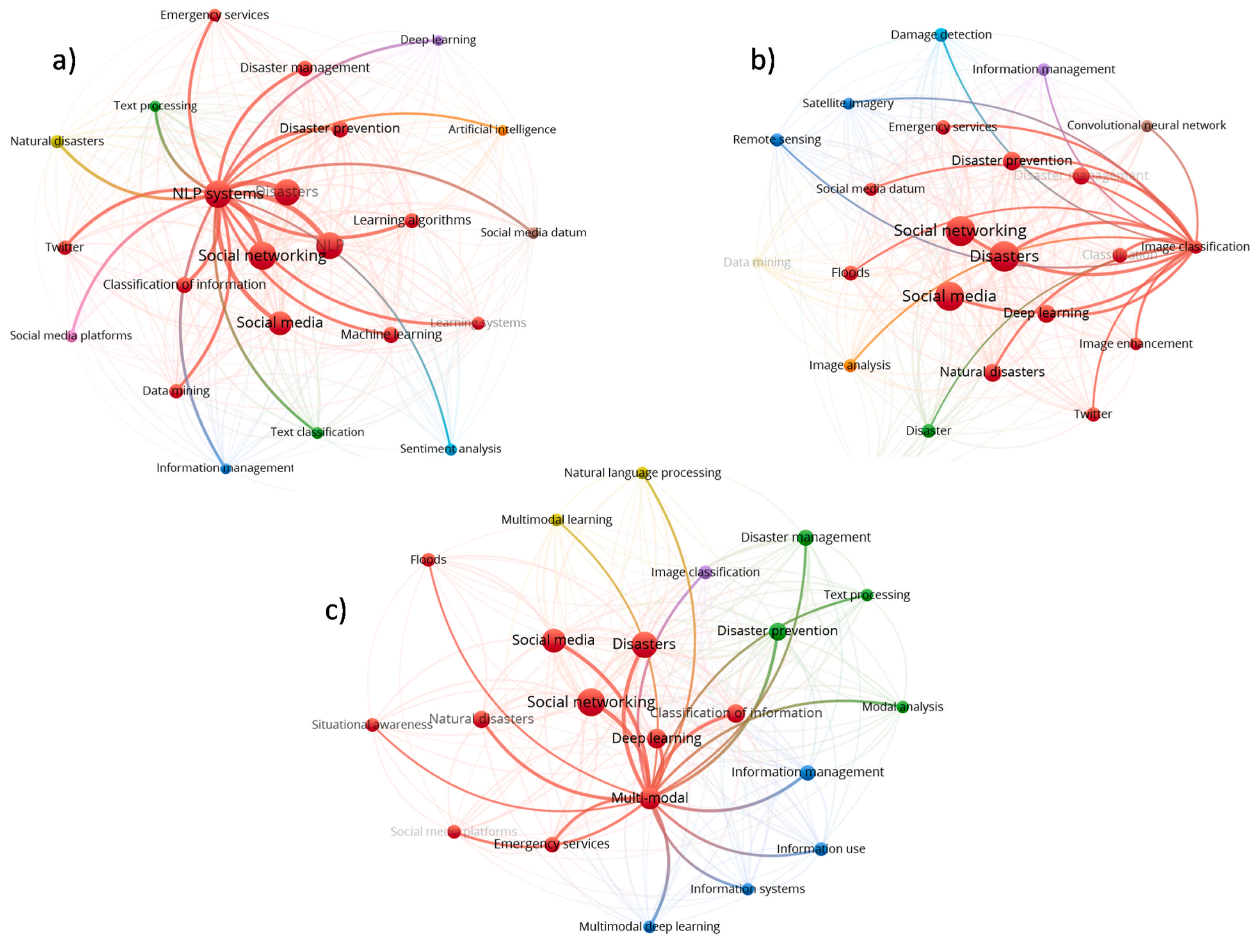

Fig. 1. Literature and research links between social media and disasters for a) natural language processing b) image processing c) multi-modal analysis.

oped ML models to detect a single dimension of information, including urgent tweets during Hurricane Harvey 2017 [36], relevant news from Times of India, NDTV India, and Indian Express [37], sentiment (positive or negative) of disaster tweets created by Figure Eight inc. (an Appen company) [38], casual sentences in newsletters during Great Hanshin-Awaji earthquake 1995 and great east Japan earthquake 2011 [39], and disaster-related tweets of Kaggle competition dataset [40] (see Table 1).

Another research stream is to extract more dimensions of humanitarian aid information from the input dataset. For instance, one of the past studies [18] classified the information dimensions into needs, availability, and other tweet classes. It has been found that multilayer perceptron models with pre-trained Word2Vec embedding have 83% classification accuracy on the test dataset [18]. Another study [41] analysed a Twitter dataset of Hurricane Harvey 2017 with annotations by topic modelling to multiple classes, including “help, donation, and recovery”, “sympathy”, “appreciation and support”, “Hurricane track and warnings”, “Damage and injuries”, and “other useful information”. “Built-environment hazards”, “Business, work, school”, “Casualty”, “Flood control infrastructure”, “Help and rescue”, “Housing”, “Preventive measures”, “Preventive measures”, “Transportation”, and “Utilities and supplies” were used as the content labels for Hurricane Harvey provided by SocialDISC [42]. These classes were utilised to develop a BiLSTM model to predict the class of tweets by Ref. [43]. Results of this study showed that 64% of tweet contents mentioned the point type or area type location.

Paul et al. [44] employed “Donation and volunteering”, “Immediate needs”, “Damage to infrastructure”, “Decreased or wounded people”, and “Unrelated or irrelevant” classes to develop CNN models based on four datasets of the Nepal earthquake 2015, Cyclone PAM 2015, California Earthquake 2014, and Typhoon Hagupit 2014 (based on CrisisNLP dataset). They demonstrated that CNN-SkipCNN (adding a few CNN layers on the top of the base CNN) has a better performance than the traditional CNN model. A victim-finder model was established to detect tweets that include “help”, “full address”, “victim info”, and “hazard situation” using several BERT models [22] and their modified versions [11]. Experimental results illustrated that BERT-based models can classify tweets better than other NLP models with limited input datasets [22].

In conclusion, according to the improvement in computational power and volume of input dataset in recent years [45] and improvement in performance accuracy of DL models relative to past ML models [46], the recent trend of studies is to serve DL models, especially BERT based on Table 1. In addition to the accuracy of classifying tweets, attention to interpretable ML models has been increasing in recent years [47]. For instance, interpretation tools, including Local Interpretable Model-Agnostic Explanation (LIME)

Table 1
Prior works on multi-modality analysis for social media posts.

N	Section 1: Textual modality			
	Input dataset	Dataset volume	Text models	
1	Manually labelled disaster tweets during the 2017 Hurricane Harvey	3191	Glove-Transformers; ELMNO-Transformers; BERT; RoBERTa; DistillBERT; ALBERT; XLNet; BERT-LSTM; BERT-CNN	
2	4 datasets based on the CrisisNLP dataset	28,647	Base CNN; CNN-GRU; CNN_SkipCNN	
3	Tweets during Hurricane Harvey 2017. Content labels by SocialDISC	4227	BiLSTM	
4	Harvard tweet dataset for hurricane Harvey 2017 - annotations by topic modelling	18,331,877	Random Forest	
5	Train: Earthquakes tweet dataset; Test: manually annotated COVID-19 tweets	122,743 + 2274	LR; CNN; MLP	
6	Kaggle competition dataset (labels are manually annotated by humans)	10,876	Trees; Random Forest; LR; Skip-gram; FastText; GloVe; Skip-gram + Bi-LSTM; FastText + Bi-LSTM; GloVe + Bi-LSTM; BERT; BERT + Bi-LSTM	
7	Manually annotated sentences from newsletters during two earthquakes	4059	SVM	
8	A twitter dataset of Figure Eight inc.	10,876	SentiBERT; BiLSTM; CNN	
9	Labelled (relevant news or not) news from Indian news	About 11,000	Multinomial Naïve; Bayes algorithm; Logistic regression; SVM; Random Forest; Xtereme Gradient Boosting Model	
10	Manually labelled tweets during Hurricane Harvey 2017	2,072,715	CNN; SVM; MLP; AdaBoost; Logistic Regression; Naïve Bayes; Decision Tree; Ridge Classifier	
N	**Section 2: Image modality**			
	Input dataset	**Dataset volume**	**Image models**	
11	Manually labelled images from Google, Twitter and BGS's image database	11,737	ResNet-50	
12	Two open-fire image datasets	1650	4 ResNets (ResNet18, 50, 101, and InceptionResNetV2)- SVM	
13	Binary annotated images from three hurricanes in 2017	7387 + 3683	VggNet; ResNet; AlexNet- SVM	
N	**Section 3: Multi-modality**			
	Input dataset	**Dataset volume**	**Image models**	**Text models**
14	CrisisMMD	12,708 (img) + 11,400 (txt)	VGG16; VGG19; ResNet50; DenseNet121; RegNetY320	DistilBERT
15	Annotated multi-labelled humanitarian tweets based on text and images (Sources: CrisisLexT26 and Datasets from Crises)	4383	DenseNet feature extraction; Faster RCNN Object detection	BERT
16	Extracted Henan torrential rain dataset in 2021	2219 (img) + 25, 880 (txt)	VGG16	BERT
18	CrisisMMD	18,126 (img) + 16,097 (txt)	VGG16	Bi-LSTM
17	CrisisMMD	18,126	VGG16; DenseNet-201	BERT
19	Annotated Twitter messenger content after Hurricane Irma 2017	6898 (image)+ 19,088(txt)	VGGNet; ResNet; Inception-V3; Tuned Inception-V3	Word2Vec
20	CrisisMMD	12,743(image)+ 11,404(txt)	VGG16	FastText
21	YFCC100 M, CrisisMMD, and tweets after Hurricane Harvey 2017 and Hurricane Irma 2017. The dataset is annotated manually.	1795 + 1555	ResNet-18 +LR; ResNet-18+decision tree; ResNet-18 +SVM with linear kernel	keyword search
22	CrisisMMD	8079 (img) +7216 (txt)	VGG16	word2vec + CNN

[18] and SHAP [48], are employed to make ML models on text data more interpretable. Results of the mentioned studies showed that interpretable ML models are capable of providing appreciated feedback such as determining the most influential words and finding the models' shortcomings.

2.2. Image modality

The second category of DL models is related to image classification to detect disasters based on social networking (as demonstrated in Fig. 1 (b)). Since DL models contain a lot of weights that should be determined in the training process, adequate training samples should be provided to train DL-based models. In addition, data collection and labelling are both costly and time-consuming [45]. To respond to these issues, pre-trained weights based on large image datasets such as ImageNet [49] are used to initialise such models [50]. For instance, pre-trained DL models were used to predict passable roads using social media and satellite images of Hurricane Harvey, Hurricane Maria, and Hurricane Irma in 2017 [51]. In addition, a labelled dataset containing 1650 images was utilised to develop four pre-trained ResNets [52] based on the ImageNet model for feature extraction and Support vector machine (SVM) for binary classification of forest fires (fire and non-fire) [53]. In another study, 11,737 labelled images were employed to develop ResNet-50 for the binary classification of landslides [54].

2.3. Multimodality

The third category, named multi-modal analysis, can consider both text and image feed to provide insightful humanitarian aid information with higher levels of classification accuracy. Most past humanitarian aid studies focused on the first category and very few focused on the third category line of research. As demonstrated in Fig. 1, mainstream multi-modal research studies currently use deep learning techniques to analyse social media feeds. The third section of Table 1 provides information about previous studies that focused on the multi-modal analysis of social media feeds. Due to powerful computers and more labelled social media posts, Table 1 shows that DL models have become more popular in recent years for both text and image modalities. ImageNet pre-trained networks including ResNet, VGG, Inception, DenseNet, and RegNet were often adopted for multi-modal analysis. BERT and its modifications were also used in the most recent multi-modality studies.

CrisisMMD [4] is one the well-known input datasets for multi-modal analysis, which consists of three steps in labelling. The first step is to find out if tweets contain informative data or not. The second step shows the humanitarian information such as “Affected individuals”, “Infrastructure and utility damage”, “Injured or dead people”, “Missing or found people”, “Rescue, volunteering, or donation effort”, “Vehicle damage”, “Other relevant information”, and “Not relevant or can't judge”. The first two steps are labelled for text and image separately. The third step is related to image modality, only determining the severity of infrastructure and utility damages. To capture all information in a pair of text and images, they should be annotated simultaneously. Concurrent labelling enables humanitarian organisations to consider the relationship between texts and images. According to the simultaneous annotation, a dataset of 4383 text image pairs data has been annotated simultaneously to find more humanitarian aid information by means of multi-modal analysis [5].

2.4. Problem statement

Earlier research on humanitarian aid has explored the use of artificial intelligence algorithms to extract insights from social media posts. These studies primarily focused on three types of input datasets: text, image, and multi-modal (combining text and image), as shown in Table 1. Presently, humanitarian aid models are centred around the multi-modal domain, offering increased value to humanitarian organisations. However, past multi-modal frameworks exhibited shortcomings in key aspects of multimodal analysis, namely input (including imbalanced datasets), model development (utilizing more accurate state-of-the-art models in contrast to earlier iterations), and the interpretation of model outputs. To address these issues, synthetic data generated through GANs are integrated to counter dataset imbalances. For the model component, advanced deep learning models like RegNetY320, ConvNeXts, BERT, and ALBERT are employed. Regarding the output, dedicated efforts are made to enhance transparency and comprehension by interpreting the results of multi-modal models.

3. Method

To answer the research questions and provide more accurate information for humanitarian aid after disasters through social media multi-modal, the DL approach is employed in this study. As demonstrated in Fig. 2, the developed framework based on multi-modal DL to analyse social media content has three main research steps, as illustrated in the following diagram and explained in the following sub-sections.

3.1. Input data collection

The first step in analysing a multi-modal model capability to provide humanitarian aid based on social media content is to prepare the datasets. The humanitarian aid datasets were annotated in Ref. [5], considering text and image contents, and are used in this study. Each pair of text and image content of a tweet was annotated based on the type of humanitarian aid, including “Caution and advice (CA)”, “Needs and offers (NO)”, “Infrastructure and utility damage (IU)”, “Affected individuals (AF)”, “Response (RE)”, “other humanitarian (OH)”, and “Not humanitarian (NH)”. Each tweet may have one or more than one humanitarian aid class that can be a source of information for humanitarian organisations after disasters. The dataset is available at: https://github.com/whuscity/multimodal-dataset-for-humanitarian-information-identification.

Fig. 3 shows the ratio of data records, with ‘detected content’ to ‘not-detected’, for each humanitarian category. If the mentioned ratio equals 0.5, it means that both detected and not-detected classes have the same number of data records. For most of the categories (except Infrastructure and Utility Damage (IU) and Response (RE)), this value is less than 0.25 which means that the number of data records in the detected class is severely less than the number of data records in the not-detected class. To address the imbalance issue of the input datasets illustrated in Fig. 3, GAN is used in this study in which new data are generated to serve the available data and to enrich classes with lower data records [55]. According to the fact that limitation on sharing images of the affected individuals is one of the main reasons for the imbalance datasets of humanitarian aid. This study only generated synthetic images to solve the mentioned issue. This synthetic data producer technique comprises a generator and a discriminator, which work against each other. New data instances are created by the generator, while the discriminator verifies the data and determines whether each instance of data is "real" from the training dataset or "fake" from the generator. The generator and discriminator are trained to work against each other until the generator can generate realistic synthetic data that the discriminator cannot classify as fake.

3.2. Multi-modal models

Pre-processing for tweet text and image is required before the multi-modal analysis is performed by DL models. Tweet text pre-processing contains unpacking contractions, converting to lowercase, and inserting space between words and punctuation marks. In addition, removing hashtag symbols (hashtag text remains), HTML character entities, text tickers, hyperlinks, whitespace, URL, RT,

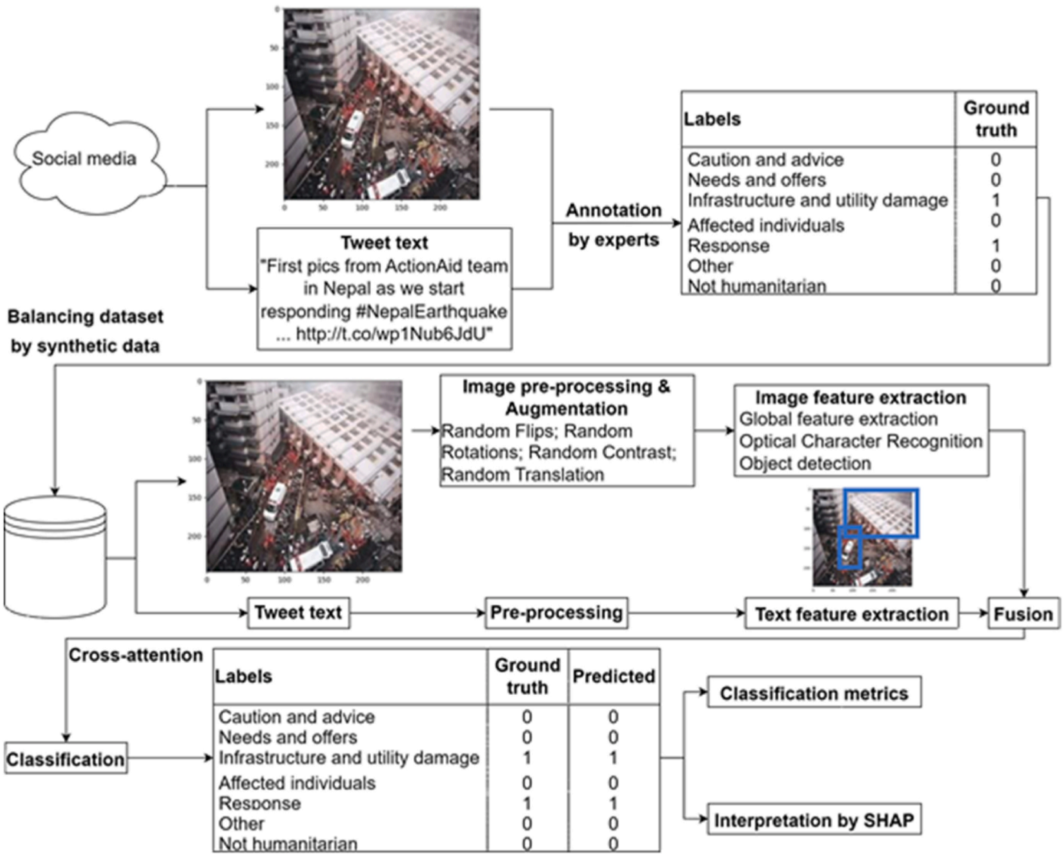

Fig. 2. The developed framework.

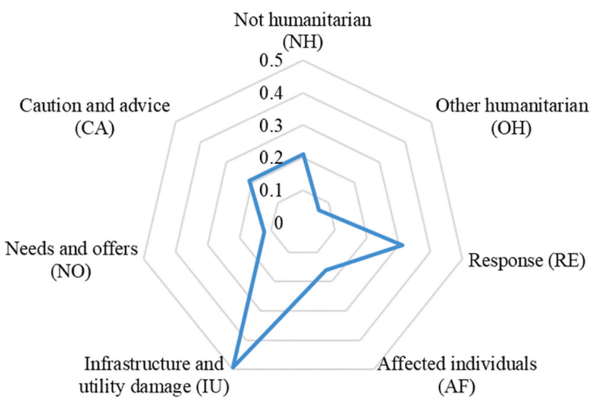

Fig. 3. Imbalanced dataset of humanitarian information in social media feeds.

mention (@), words with two or fewer letters, symbols (&, < and >), misspelling words, punctuation, emoji, Basic Multilingual Plane (BMP) character, stop words are done in the text pre-processing. For tweet images, image augmentation is done to increase the number of training images by moving the image along the X or Y direction (or both), rotating an image about a horizontal or vertical axis, and changing the lightness and darkness of images.

Before extracting global features from a pair of text and image, Optical Character Recognition (OCR) and object detection algorithms are employed to analyse the tweet image. OCR is utilised in this study to analyse tweet images and extract text from them [56]. OCR and object detection are two common approaches that researchers use to analyse social media content [57]. In addition, object detection is employed to detect objects in an image, and determine the position of the detected objects [58]. A Faster-RCNN model [59] is selected for object detection in this framework. ImageNet pre-trained Faster-RCNN Inception Resnet V2 [60] that is based on Resnet and Inception Resnet feature extractors is conducted in this study.

For global feature extraction, the original BERT [22] and its modified versions (ALBERT) [23] are selected for textual modality. BERT and ALBERT are chosen based on their outstanding performance on social media feeds to provide aid information after natural disasters [11]. For image modality, two recent ImageNet pre-trained deep CNNs (RegNetY320 [24] and ConvNeXts [25]) are considered. RegNetY320 is a simple and fast network that is employed due to its acceptable performance for CrisisMMD in classifying humanitarian aid information. Moreover, ConvNeXts are built from standard ConvNet modules and have better performance in a variety of image applications [26]. TensorFlow Hub is used to develop BERT and ALBERT, while the Keras application programming interface is employed for RegNetY320 and ConvNeXts.

The early fusion strategy integrates the extracted multi-modality features. Early fusion provides more flexibility for multimodality analysis [61] and demonstrates better performance than late fusion for emotion classification [62]. According to the suggestion of past studies [63], the cross-attention mechanism is performed after fusion to concentrate on the important part of the social media feeds. Cross-attention is a novel module to determine the target object regions and improve the discrimination of the extracted feature [64]. Due to the imbalanced datasets, focal loss [65] is utilised as the loss function in the developed framework. Focal loss is a variant of the binary cross-entropy loss, and it is employed in this study. Focal loss is a powerful solution to improve the performance of classifiers for imbalanced datasets [66,67]. The equation of focal loss is mentioned below [65,68]:

$$FL\left(p_t\right) = -\alpha_t\left(1-p_t\right)^{\gamma}\log\left(p_t\right) \tag{1}$$

where γ is the focusing parameter on the imbalance class; if $\gamma = 0$ focal loss will be similar to cross-entropy; α_t is the balancing ratio; p_t can be evaluated based on the following equation:

$$p_t = \begin{cases} -p, y = 1 \\ 1-p, otherwise, \end{cases} \tag{2}$$

where y is the ground-truth class (one class of humanitarian aid information); p is the deep learning model estimated probability. Python and relevant packages are employed for generating and verifying all models. The parameter settings of the developed framework based on trial and error are mentioned in Table 2. To clarify the process, the following trial and error is designed for efficient tuning of the developed framework and the results are demonstrated in Fig. 4. In Fig. 4 (parallel coordinates plot), the concept explored involves assessing the performance of a developed framework under different conditions. This assessment involves experimenting with various configurations, such as different batch sizes (16 and 32), diverse dropout rates (0.2, 0.25, and 0.3), and two optimizer algorithms: Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam). A total of 12 experiments are conducted, with results visualized through red lines indicating higher accuracy and blue lines indicating lower accuracy. Upon compar-

Table 2
Parameter settings of the developed framework.

Parameter	Setting
Input image size	224 × 224 pixels
Dropout rate	0.25
Activation function	ReLU, Sigmoid
Batch size	32
Learning rate	0.00002
Optimizer	Adam
Loss function	Binary Focal Loss (gamma = 2)
Number of hidden layers after fusion	2
Number of neurons in each hidden layer after fusion	256, 64
Number of GAN synthetic images for each run	25

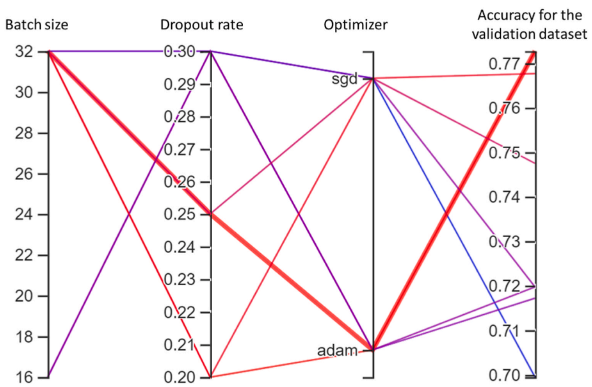

Fig. 4. Setting of the hyper-parameters of deep learning models.

ing the outcomes of these experiments, specific values emerge as optimal choices. Specifically, a batch size of 32, a dropout rate of 0.25, and the Adam optimizer yield the best performance. The model's accuracy for the chosen hyperparameters on the validation dataset is 77.27%. In essence, this analysis explores the impact of various parameters on the accuracy of the developed framework, considering different batch sizes, dropout rates, and optimizer algorithms.

3.3. Performance evaluation and interpretation

Accuracy, precision, recall, and F1-score are employed to evaluate the performance of the classifier models in this study. Eqs. (3)–(6) show how to calculate accuracy, precision, recall and F1-score, respectively.

$$\text{Accuracy} = \frac{\textit{True positive} + \textit{True negative}}{\textit{True positive} + \textit{True negative} + \textit{False positive} + \textit{Flase negative}} \tag{3}$$

$$\text{Precision} = \frac{\textit{True positive}}{\textit{True positive} + \textit{Flase positive}} \tag{4}$$

$$\text{Recall} = \frac{\textit{True positive}}{\textit{True positive} + \textit{Flase negative}} \tag{5}$$

$$\text{F1} - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

True positive (ground truth and predicted = 1) and true negative (ground truth and predicted = 0) are correct classifications of humanitarian aid information. A false positive (ground truth = 1; and predicted = 0) and false negative (ground truth = 0; and predicted = 1) mean that the ML model fails to accurately predict the class.

In this study, SHAP with its Python library [69] was selected and implemented to understand the behaviour of the developed framework in classifying humanitarian aid social feed. Shapely values can demonstrate the contribution of each word in classification and detect the importance of humanitarian aid information. The selection of SHAP for interpretation of the deep learning for humanitarian aid models in this study is related to three key attributes of the SHAP. First, SHAP analysis considers the interdependencies among variables to quantify the weight of each feature in the ML model [70]. Second, SHAP is able to fairly distribute the importance of each feature that can be important for humanitarian aid activities [71]. Third, SHAP is the most comprehensive method for explaining machine learning models, and it can be used for different types of the data such as text and images [72]. It is worth noting that each modality is interpreted separately using shapely values.

4. Experimental results

The collected dataset was classified by using the developed framework, as demonstrated in Fig. 2. The objective of the framework is to classify pairs of text and images correctly into their respective classes. Accurate classification will provide a piece of valuable humanitarian aid information for humanitarian organisations to take action after natural disasters. In the training phase of all networks, the collected dataset was randomly divided into the training, validation, and testing datasets in a 0.85:0.05:0.1 ratio in a stratified fashion. The comparison among developed models is done using unseen test datasets. Experiments are performed with an Intel Core i7-11700 @ 2.5 GHz, 32 GB and NVIDIA Quadro P620. The experimental results can be divided into three categories, as presented below.

4.1. Model performance comparison

Evaluation metrics of the developed framework based on various text (BERT and ALBERT) and image (RegNets and ConvNeXts) feature extractions are presented in Table 3. The “WA” is a weighted average of performance metrics of all categories. Comparing the first two models in Tables 3 and it is found that the F1 scores of the first row (BERT + RegNetY320) are higher than those in the second row (ALBERT + RegNetY320). The F1 scores for the two first models from the AF category are the same. In addition, the weighted average of all performance metrics of the first model is higher than that of the second one. Since the difference in the first two models is related to the text feature extraction model, it is found that the performance of text feature extraction of the first model (BERT) is better than the second model (ALBERT). Similar results can also be seen from the comparison of the last two models in Table 3. The F1 scores of the BERT feature extraction model are higher than those of the ALBERT model for four categories. For two categories including NO and AF, the F1-scores of ALBERT are higher than those of BERT. For the NH (Not humanitarian) category, the F1 scores of the two last models are the same. Weighted averages of all performance metrics of the BERT model (third model) are higher than that of the ALBERT (fourth model).

To compare the performance of image feature extraction models (RegNet and ConvNeXt), the first and third models should be compared with each other. For all seven categories, the performance of the third model (ConvNeXt) is better than the first model (RegNet). Furthermore, the comparison between the second and fourth models also proves the above-mentioned statement. Based on the outstanding performance of BERT and ConvNeXt in comparison with ALBERT and RegNet respectively, the performance of the third model (i.e., BERT + ConvNeXtXLarge) is better than the other models in Table 3. The accuracy, precision, recall, and F1-score of this model range from 0.816 to 0.923, 0.773 to 0.927, 0.814 to 0.922, and 0.749 to 0.920, respectively. On the other hand, the second model in Table 3 (ALBERT and RegNet) has the lowest classifying performance in comparison with other models. For instance, the F1-score for detecting IU (Infrastructure and Utility damage) category is 0.656 which is the lowest score in Table 3. Training time

Table 3
Comparison among the prediction performance of four feature extraction models based on the test dataset for each category and the weighted average (WA).

		CA	NO	IU	AF	RE	OH	NH	WA	Time (s)
1	*BERT + RegNetY320*									*2418*
	• *Accuracy*	0.893	*0.904*	*0.811*	*0.836*	*0.813*	***0.940***	*0.813*	*0.841*	
	• *Precision*	*0.880*	*0.881*	*0.803*	*0.702*	*0.810*	***0.944***	*0.760*	*0.811*	
	• Recall	*0.891*	*0.902*	*0.811*	***0.836***	*0.811*	***0.939***	*0.813*	*0.840*	
	• *F1-score*	*0.884*	*0.874*	*0.804*	*0.761*	*0.785*	*0.915*	*0.745*	*0.810*	
2	*ALBERT + RegNetY320*									*1588*
	• *Accuracy*	*0.852*	*0.895*	*0.729*	*0.836*	*0.772*	*0.934*	*0.813*	*0.803*	
	• *Precision*	*0.839*	*0.835*	*0.733*	*0.702*	*0.749*	*0.869*	*0.752*	*0.766*	
	• *Recall*	*0.852*	*0.893*	*0.730*	***0.836***	*0.770*	*0.934*	*0.815*	*0.803*	
	• *F1-score*	*0.810*	*0.850*	*0.656*	*0.761*	*0.739*	*0.906*	*0.736*	*0.746*	
3	*BERT + ConvNeXtXLarge*									*6850*
	• *Accuracy*	***0.923***	***0.911***	***0.868***	*0.829*	***0.827***	*0.934*	*0.816*	***0.863***	
	• *Precision*	***0.927***	*0.897*	***0.869***	*0.781*	***0.819***	*0.911*	*0.773*	***0.848***	
	• *Recall*	***0.922***	*0.907*	***0.865***	*0.831*	***0.825***	*0.931*	*0.814*	***0.861***	
	• *F1-score*	***0.920***	*0.881*	***0.867***	*0.783*	***0.818***	***0.917***	***0.749***	***0.842***	
4	*ALBERT + ConvNeXtXLarge*									*5924*
	• *Accuracy*	*0.911*	***0.911***	*0.863*	***0.838***	*0.795*	*0.932*	***0.820***	*0.855*	
	• *Precision*	*0.902*	***0.900***	*0.862*	***0.800***	*0.788*	*0.869*	***0.816***	*0.843*	
	• *Recall*	*0.909*	***0.912***	*0.861*	***0.836***	*0.798*	*0.934*	***0.823***	*0.855*	
	• *F1-score*	*0.907*	***0.902***	*0.862*	***0.800***	*0.788*	*0.900*	***0.749***	*0.836*	

of ConvNeXt is much higher than that of RegNet; however, the training time of ALBERT and BERT is approximately similar (As mentioned in Table 3). Finally, it can be concluded that the third model in Table 3 (i.e., BERT + ConvNeXtXLarge) is selected as the most appropriate DL architecture to analyse crowd-sourced data related to natural disasters in remote regions. Consequently, the answer to the first research question is determined. Humanitarian aid organisations can use the most accurate models (BERT + ConvNeXtXLarge) to predict the category of future crowd-sourced data after a disaster and provide help in a timely manner. Using (BERT + ConvNeXtXLarge) helps practitioners to find more information from social media content after disasters.

4.2. GAN dataset augmentation impact

A comparison of the performance metrics including accuracy, precision, recall, and F1-score of BERT and ConvNeXtXLarge models with and without GAN images in the model development process is illustrated in Fig. 5 (a) for the test dataset. The comparison is made between the use of BERT and ConvNeXtXLarge models based on their results (see Section 4.1). CA, NO, and AF are selected for this comparison due to the highly imbalanced classes (Fig. 3). Fig. 5 (a) demonstrates that synthetic images generated by GAN approximately improve the performance metrics of the models. It is worth noting that synthetic images are not readable for humans, however, based on the results deep learning models can recognize them.

Fig. 5 (b) shows the confusion matrices of the classifiers to clarify the impact of GAN augmentation datasets on the models. As demonstrated in Fig. 5 (b), the number of detected class and not-detected class that is accurately classified increases for all three categories (CA, NO, and AF) with GAN. This means that GAN provides more data for the model learning process, and the models can predict this category better than they could in the past. Consequently, the second research question answers that GAN is the recommended approach to improve the model performance for classifying imbalanced datasets.

4.3. Interpretation of results with SHAP

The results of the SHAP analysis, which provides valuable insights for humanitarian organisations, are given in Fig. 6. Fig. 6 interprets the image and text of three tweets that are related to IU, AF, and RE. The important parts of tweet images are highlighted in red. Furthermore, Fig. 6 shows that "damage" and "road" are the most influential words in the tweet texts to detect the IU. For the AF category, "toll", "missing", "Britons", "Nepal", and "rises" are the most important words that have a positive impact. Finally, "relief" and "army" are the most influential words in the tweet texts to detect the RE. Most of the results are similar to those expected, and it shows that the model is working properly. Consequently, this interpretation could be monitored in the beginning steps of natural disasters to explain the model behaviour.

In order to compare the results of the SHAP with another tool, a class activation map (CAM) is employed in this study as a classic tool for an explanation of deep learning models. CAM is able to demonstrate the most important regions of the images in the classification tasks [73]. After evaluating the two interpretation methods for images including SHAP and CAM (Fig. 6), it is obvious that CAM results make sense more for readers. Especially for humanitarian aid images that have various contexts. In addition, SHAP privilege in comparison with CAM is related to its ability to interpret text in addition to images.

5. Discussion

Humanitarian aid in post-disaster situations should be provided in a timely manner. However, the affected people and society lack the ability to express their needs [27]. To solve this issue, recent studies suggested [5,11] analysing the crowd-sourced data of the affected people. The affected people can express their needs by using social media platforms. If humanitarian organisations are able to

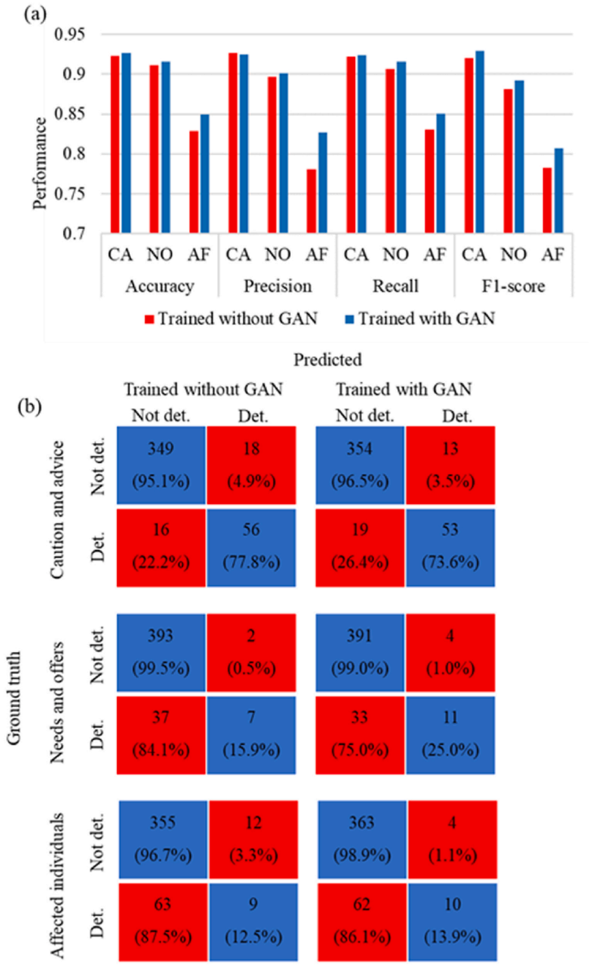

Fig. 5. Comparison of the performance of BERT and ConvNeXtXLarge model for the test datasets included and excluded GAN images in the training process (a) Bar chart (b) Confusion matrix.

classify the crowd-sourced data and find informative content, they will be capable to deliver essential items such as medicine and food as soon as possible [74]. Accurate classifying crowd-sourced data is a critical and lifesaving task, so it should be done with accurate models.

Due to the increase in humanitarian aid demand [75], humanitarian organisations should enhance the efficiency of humanitarian aid [76]. Thus, deep learning algorithms are preferred to traditional machine learning algorithms in this study. Previous studies have demonstrated that the performance of deep networks is superior to that of traditional models, provided that a large and uniformly distributed input dataset is accessible. In addition, crowd-sourced data has various modalities such as text and images that can be used to capture more critical information from informative content. In a post-disaster scenario, information sourced from various modalities holds significant value for humanitarian aid efforts [77]. Employing a multimodal approach allows us to harness a broader spectrum of information. Moreover, by embracing a multimodal perspective, we can effectively explore the interconnectedness among different modalities. This exploration results in the generation of more meaningful and context-rich data that proves invaluable to humanitarian aid organisations. Therefore, the practice of multimodal analysis aligns seamlessly with a prevailing trend in contemporary literature. This trend involves the integration of diverse modalities into advanced models, such as KOSMOS [78], Video-LLaMA [79], and Flamingo [80], all of which have gained prominence due to their capability to address multifaceted challenges effectively. Consequently, this study analyses imbalanced input datasets to develop an interpretable multi-modal framework. The following sections discuss the comparison of utilised deep networks, GAN as a solution for dealing with the imbalanced dataset, and the interpretation of deep networks for a better understanding of their behaviour, in response to the research questions mentioned in the introduction.

5.1. Deep neural networks comparison

Previous studies have used several pre-trained deep learning models, including VGG, ResNet, Inception-V3, DenseNet, and RegNet in the multi-modal analysis. It was reported that pre-trained deep networks have the appropriate performance to extract features of social media images [27]. This research employs RegNet and ConvNeXt to evaluate their applicability for multimodality analysis of social media feeds. Experimental results show that ConvNeXt performance has the best performance compared to RegNet. The useful-

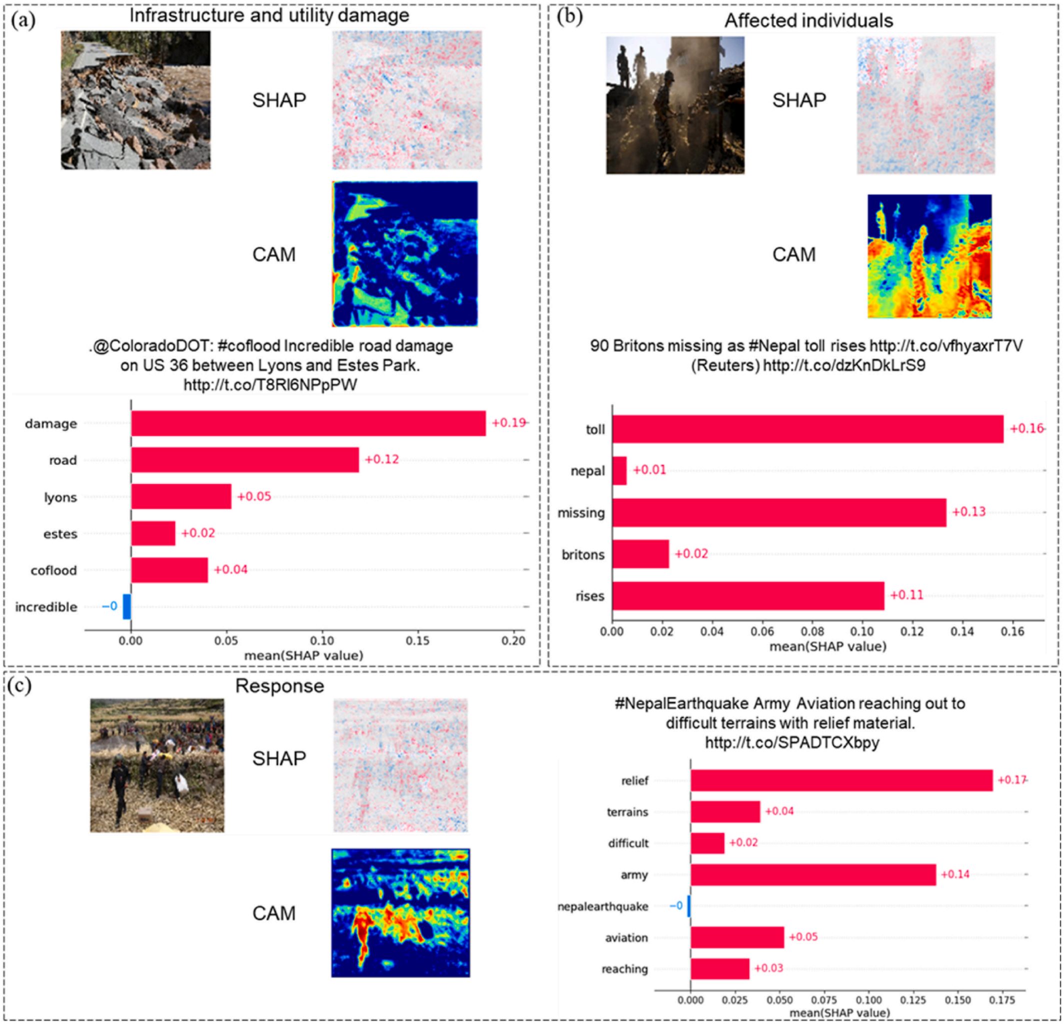

Fig. 6. Interpretation of results by SHAP to detect humanitarian aid data including (a) Infrastructure and utility damage (b) Affected individuals (c) Response.

ness of ConvNeXt is in line with the results of past studies that employed it for medical images [26,81], waste classification [82], and detecting seismic damages to buildings [83].

For the text modality, the capability of deep networks in feature extraction is higher than that of the traditional ML model as reported in the literature. Consequently, this study uses BERT and one of its modifications (ALBERT) to analyse the text inputs of the social media feeds. The ability of BERT to analyse social media feeds has been investigated and proven by past studies [6]. Results show that the BERT model has better performance than ALBERT in this study. These observations are consistent with those in the literature that compared numerous NLP models and found that BERT is better than some of its modifications including ALBERT, DistilBert, and RoBERTa [11]. As a result, ConvNeXts and BERT are selected as the appropriate DL architectures to analyse crowd-sourced data related to natural disasters in remote regions (RQ1). Practitioners can use the multi-modal framework to predict various information of social media content in the future and provide humanitarian aid as soon as possible.

5.2. GAN performance

Due to a few reasons such as privacy policy (sharing the images of affected individuals is prevented) [84], social media feeds tend to be highly imbalanced [16]. The possibility of GAN to solve this issue is analysed in this study. In addition, collecting and labelling real data records are timely and costly expensive [45]. Therefore, synthetic data records can be useful in improving accuracy when access to real data records is limited. Experimental results proved that synthetic GAN images can increase performance metrics by reducing data imbalance. The importance of this improvement can be highlighted by the fact that accurate detection of humanitarian aid tweets can save human lives. The usefulness of synthetic images by GAN for multi-modal analysis of crowd-sourced data is consistent with the findings of past research studies that used GAN for other applications such as brain tumour MRI image [85,86], machine fault diagnosis [87], corrosion [88] and structure health monitoring. Consequently, the accuracy of the model for humanitarian aid data from social media can be increased by generating synthetic data records.

5.3. Interpretation of deep networks

As mentioned in Section 5.1, the performance of the DL makes them suitable for classifying humanitarian information from social media feeds. However, the complexity of DL models also means that it is hard to interpret and explain the behaviour. According to the seriousness of the humanitarian aid issue, utilizing black-box models is not rational [89]. This study employed SHAP to interpret and explain DL results. The interpretation can be used in the initial stages of DL as a sanity check to validate the results [18]. Consequently, the interpretation results can enhance the real-world applications of DL for social media feeds. The usefulness of the interpretation tools for deep networks is demonstrated, in line with the findings of past studies that interpret the textual modality of crowd-sourced data [18], COVID-19 claims [90], and linguistic features of Italian social media [91]. Thus, humanitarian aid experts can see the behaviour of deep learning models before usage in real disasters as a sanity check.

6. Conclusion

This study has developed an interpretable multimodal framework that can effectively classify crowd-sourced data in social media feeds with high accuracy. The developed framework contains text and image pre-processing and optical character recognition (OCR). In addition, several pre-trained deep neural networks including BERT, ALBERT, ConvNeXt and RegNet, and Faster RCNN are employed in the framework. BERT and ALBERT are related to text modality, while other deep networks are devoted to image modality. The experimental results for the test datasets illustrate that the average F1-score and accuracy of the combination of ConvNeXt and BERT are 0.842 and 0.863, respectively, and its performance is better than that of the other combinations. Since social media feeds are inclined to imbalance, the applicability of generative adversarial networks (GAN) to solve data imbalance issues is demonstrated by applying GAN to the developed framework. The results prove that using synthetic data records in training can increase the F1-score and accuracy of the developed model for the test datasets by roughly 2%. Finally, it is shown that SHAP is a suitable tool to interpret the complex behaviour of deep networks.

Data limitation is one of the main limitations of the current study. All experiments of this research study are performed on public social media feeds for several types of disasters. Since classification accuracy densely depends on the types of disasters, it is recommended that future studies develop multi-modal datasets for the specific types of natural disasters including wildfires, floods, and earthquakes. Moreover, due to the usefulness of synthetic data records in solving the imbalanced issue, it is suggested that future studies find the optimal number of synthetic data records to improve classification performance. For a more profound enhancement of the quality and effectiveness of synthetic data records, the presence of a significant dataset becomes essential. This is particularly crucial when aiming to generate a wide array of diverse image variations to create synthetic content of superior quality. Furthermore, it is suggested that future studies evaluate the impact of multi-modal synthetic data (generating both text and image) on the prediction performance of deep learning models. Additionally, the modalities are interpreted in this study separately, it is recommended that future studies interpret the multi-models as a whole and compare the results with this study. Interpretation of the correlation between two modalities can be a potential topic for future research studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors are grateful for support from the Australian Research Council (ARC) through the Linkage project (LP180101080). The authors acknowledge the contributions of the members of ASCII Lab at Monash University for critiquing the manuscript and providing constructive feedback.

References

[1] X. Ning, L. Yao, X. Wang, B. Benatallah, Calling for response: automatically distinguishing situation-aware tweets during crises, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10604, LNAI, 2017, pp. 195–208 https://doi.org/10.1007/978-3-319-69179-4_14 [Online]. Available:

[2] L. Li, R. Sampson, S.H. Ding, L. Song, TASR: adversarial learning of topic-agnostic stylometric representations for informed crisis response through social media, Inf. Process. Manag. 59 (2) (2022) 102857.

[3] S. Shan, F. Zhao, Y. Wei, Real-time assessment of human loss in disasters based on social media mining and the truth discovery algorithm, Int. J. Disaster Risk Reduc. 62 (2021), 102418 https://doi.org/10.1016/j.ijdrr.2021.102418, 2021/08/01/.

[4] F. Alam, F. Ofli, M. Imran, Crisismmd: multimodal twitter datasets from natural disasters, in: Twelfth International AAAI Conference on Web and Social Media, 2018.

[5] X. Wu, J. Mao, H. Xie, G. Li, Identifying humanitarian information for emergency response by modeling the correlation and independence between text and images, Inf. Process. Manag. 59 (4) (2022) 102977.

[6] R. Suwaileh, T. Elsayed, M. Imran, H. Sajjad, When a disaster happens, we are ready: location mention recognition from crisis tweets, Int. J. Disaster Risk Reduc. 78 (2022), 103107 https://doi.org/10.1016/j.ijdrr.2022.103107, 2022/08/01/.

[7] R. Grace, Overcoming barriers to social media use through multisensor integration in emergency management systems, Int. J. Disaster Risk Reduc. 66 (2021), 102636 https://doi.org/10.1016/j.ijdrr.2021.102636, 2021/12/01/.

[8] C. Rajeshkannan, S. Kogilavani, Modelling of flood prediction by optimizing multimodal data using regression network, in: Mobile Computing and Sustainable

Informatics, Springer, 2022, pp. 489–511.
[9] J. Tang, S. Yang, W. Wang, Social media-based disaster research: development, trends, and obstacles, Int. J. Disaster Risk Reduc. 55 (2021) 102095.
[10] M. Arashpour, V. Kamat, A. Heidarpour, M.R. Hosseini, P. Gill, Computer vision for anatomical analysis of equipment in civil infrastructure projects: theorizing the development of regression-based deep neural networks, Autom. ConStruct. 137 (2022), 104193 https://doi.org/10.1016/j.autcon.2022.104193, 2022/05/01/.
[11] B. Zhou, et al., VictimFinder: harvesting rescue requests in disaster response from social media with BERT, Comput. Environ. Urban Syst. 95 (2022), 101824 https://doi.org/10.1016/j.compenvurbsys.2022.101824, 2022/07/01/.
[12] F. Ofli, F. Alam, M. Imran, Analysis Of Social Media Data Using Multimodal Deep Learning For Disaster Response, 2020 arXiv preprint arXiv:2004.11838.
[13] G. Bhullar, A. Khullar, A. Kumar, A. Sharma, H.S. Pannu, A. Malhi, Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient resource management, Int. J. Disaster Risk Reduc. 75 (2022), 102979 https://doi.org/10.1016/j.ijdrr.2022.102979, 2022/06/01/.
[14] C. Rossi, et al., Early detection and information extraction for weather-induced floods using social media streams, Int. J. Disaster Risk Reduc. 30 (2018) 145–157 https://doi.org/10.1016/j.ijdrr.2018.03.002, 2018/09/01/.
[15] M. Buda, A. Maki, M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Network. 106 (2018) 249–259 https://doi.org/10.1016/j.neunet.2018.07.011, 2018/10/01/.
[16] X. Tao, et al., SVDD-based weighted oversampling technique for imbalanced and overlapped dataset learning, Inf. Sci. 588 (2022) 13–51 https://doi.org/10.1016/j.ins.2021.12.066, 2022/04/01/.
[17] M. Arashpour, T. Ngo, H. Li, Scene understanding in construction and buildings using image processing methods: a comprehensive review and a case study, J. Build. Eng. 33 (2021), 101672 https://doi.org/10.1016/j.jobe.2020.101672, 2021/01/01/.
[18] S. Behl, A. Rao, S. Aggarwal, S. Chadha, H.S. Pannu, Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises, Int. J. Disaster Risk Reduc. 55 (2021), 102101 https://doi.org/10.1016/j.ijdrr.2021.102101, 2021/03/01/.
[19] Z. Zou, H. Gan, Q. Huang, T. Cai, K. Cao, Disaster image classification by fusing multimodal social media data, ISPRS Int. J. Geo-Inf. 10 (10) (2021), https://doi.org/10.3390/ijgi10100636.
[20] S.D. Mohanty, et al., A multi-modal approach towards mining social media data during natural disasters-A case study of Hurricane Irma, Int. J. Disaster Risk Reduc. 54 (2021) 102032.
[21] H. Hao, Y. Wang, Leveraging multimodal social media data for rapid disaster damage assessment, Int. J. Disaster Risk Reduc. 51 (2020) 101760.
[22] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-Training Of Deep Bidirectional Transformers For Language Understanding, 2018 arXiv preprint arXiv:1810.04805.
[23] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A Lite Bert For Self-Supervised Learning Of Language Representations, 2019 arXiv preprint arXiv:1909.11942.
[24] I. Radosavovic, R.P. Kosaraju, R. Girshick, K. He, P. Dollár, Designing network design spaces, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10428–10436.
[25] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.
[26] Z. Han, M. Jian, G.-G. Wang, ConvUNeXt: an efficient convolution neural network for medical image segmentation, Knowl. Base Syst. 253 (2022), 109512 https://doi.org/10.1016/j.knosys.2022.109512, 2022/10/11/.
[27] S. Kotha, S. Haridasan, A. Rattani, A. Bowen, G. Rimmington, and A. Dutta, "Multimodal combination of text and image tweets for disaster response assessment," 2022: International Workshop on Data-Driven Resilience Research..
[28] B. Wu, L. Liu, Y. Yang, K. Zheng, X. Wang, Using improved conditional generative adversarial networks to detect social bots on Twitter, IEEE Access 8 (2020) 36664–36680.
[29] T. Zhou, Q. Li, H. Lu, Q. Cheng, X. Zhang, GAN review: models and medical image fusion applications, Inf. Fusion 91 (2023/03/01/2023) 134–148, https://doi.org/10.1016/j.inffus.2022.10.017.
[30] M. Haggag, A. Yosri, W. El-Dakhakhni, E. Hassini, Interpretable data-driven model for Climate-Induced Disaster damage prediction: the first step in community resilience planning, Int. J. Disaster Risk Reduc. 73 (2022), 102884 https://doi.org/10.1016/j.ijdrr.2022.102884, 2022/04/15/.
[31] M. Arashpour, et al., Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization, Comput. Appl. Eng. Educ. 31 (1) (2023) 83–99, https://doi.org/10.1002/cae.22572.
[32] Y. Chen, S. He, Z. Zhou, Investigation of social media representation bias in disasters: towards a systematic framework, Int. J. Disaster Risk Reduc. 81 (2022), 103312 https://doi.org/10.1016/j.ijdrr.2022.103312, 2022/10/15/.
[33] E. Ntontis, et al., Tracking the nature and trajectory of social support in Facebook mutual aid groups during the COVID-19 pandemic, Int. J. Disaster Risk Reduc. 76 (2022), 103043 https://doi.org/10.1016/j.ijdrr.2022.103043, 2022/06/15/.
[34] R.I. Ogie, S. James, A. Moore, T. Dilworth, M. Amirghasemi, J. Whittaker, Social media use in disaster recovery: a systematic literature review, Int. J. Disaster Risk Reduc. 70 (2022), 102783 https://doi.org/10.1016/j.ijdrr.2022.102783, 2022/02/15/.
[35] S. Guha, R.K. Jana, M.K. Sanyal, Artificial neural network approaches for disaster management: a literature review, Int. J. Disaster Risk Reduc. 81 (2022), 103276 https://doi.org/10.1016/j.ijdrr.2022.103276, 2022/10/15/.
[36] A. Devaraj, D. Murthy, A. Dontula, Machine-learning methods for identifying social media-based requests for urgent help during hurricanes, Int. J. Disaster Risk Reduc. 51 (2020), 101757 https://doi.org/10.1016/j.ijdrr.2020.101757, 2020/12/01/.
[37] J. Domala, et al., Automated identification of disaster news for crisis management using machine learning and Natural Language Processing, Available: in: Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC, 2020, pp. 503–508, https://doi.org/10.1109/ICESC48915.2020.9156031, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090839794&doi=10.1109%2fICESC48915.2020.9156031&partnerID=40&md5=7b1a7fe9c331bdbad7a7189e2c21fd6f.
[38] G. Song, D. Huang, A sentiment-aware contextual model for real-time disaster prediction using twitter data, Future Internet 13 (7) (2021), https://doi.org/10.3390/fi13070163.
[39] F. Sakahira, U. Hiroi, Designing cascading disaster networks by means of natural language processing, Int. J. Disaster Risk Reduc. 66 (2021) 102623.
[40] S. Deb, A.K. Chanda, Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data, Mach. Learn. Appl. 7 (2022/03/15/2022) 100253, https://doi.org/10.1016/j.mlwa.2022.100253.
[41] M. Karimiziarani, K. Jafarzadegan, P. Abbaszadeh, W. Shao, H. Moradkhani, Hazard risk awareness and disaster management: extracting the information content of twitter data, Sustain. Cities Soc. 77 (2022), 103577 https://doi.org/10.1016/j.scs.2021.103577, 2022/02/01/.
[42] C. Zhang, W. Yao, Y. Yang, R. Huang, A. Mostafavi, Semiautomated social media analytics for sensing societal impacts due to community disruptions during disasters, Comput. Aided Civ. Infrastruct. Eng. 35 (12) (2020) 1331–1348.
[43] A.S. Paradkar, C. Zhang, F. Yuan, A. Mostafavi, Examining the consistency between geo-coordinates and content-mentioned locations in tweets for disaster situational awareness: a Hurricane Harvey study, Int. J. Disaster Risk Reduc. 73 (2022), 102878 https://doi.org/10.1016/j.ijdrr.2022.102878, 2022/04/15/.
[44] N.R. Paul, D. Sahoo, R.C. Balabantaray, Classification of crisis-related data on Twitter using a deep learning-based framework, Article. Multimedia Tools Appl. (2022), https://doi.org/10.1007/s11042-022-12183-w.
[45] Z. Zhang, N. He, D. Li, H. Gao, T. Gao, C. Zhou, Federated transfer learning for disaster classification in social computing networks, J. Saf. Sci. Resilience 3 (1) (2022) 15–23 https://doi.org/10.1016/j.jnlssr.2021.10.007, 2022/03/01/.
[46] P.K. Roy, A. Kumar, J.P. Singh, Y.K. Dwivedi, N.P. Rana, R. Raman, Disaster related social media content processing for sustainable cities, Sustain. Cities Soc. 75 (2021), 103363 https://doi.org/10.1016/j.scs.2021.103363, 2021/12/01/.
[47] M. Arashpour, AI explainability framework for environmental management research, J. Environ. Manag. 342 (2023), 118149 https://doi.org/10.1016/j.jenvman.2023.118149, 2023/09/15/.
[48] J.G.D. Harb, R. Ebeling, K. Becker, A framework to analyze the emotional reactions to mass violent events on Twitter and influential factors, Inf. Process. Manag. 57 (6) (2020) 102372, https://doi.org/10.1016/j.ipm.2020.102372.

[49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A Large-Scale Hierarchical Image Database, 2009, 2009, pp. 248–255 Ieee.
[50] Z. Fang, J. Ren, S. Marshall, H. Zhao, S. Wang, X. Li, Topological optimization of the DenseNet with pretrained-weights inheritance and genetic channel selection, Pattern Recogn. 109 (2021), 107608 https://doi.org/10.1016/j.patcog.2020.107608, 2021/01/01/.
[51] K. Ahmad, et al., Automatic detection of passable roads after floods in remote sensed and social media data, Signal Process. Image Commun. 74 (2019/05/01/ 2019) 110–118, https://doi.org/10.1016/j.image.2019.02.002.
[52] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2016, pp. 770–778 2016.
[53] S. Dogan, et al., Automated accurate fire detection system using ensemble pretrained residual network (in English). Expert Syst. Appl. 203 (2022) 117407, https://doi.org/10.1016/j.eswa.2022.117407.
[54] C.V. Pennington, et al., A near-real-time global landslide incident reporting tool demonstrator using social media and artificial intelligence, Int. J. Disaster Risk Reduc. 77 (2022) 103089.
[55] S. Liu, H. Jiang, Z. Wu, Y. Liu, K. Zhu, Machine fault diagnosis with small sample based on variational information constrained generative adversarial network, Adv. Eng. Inf. 54 (2022), 101762 https://doi.org/10.1016/j.aei.2022.101762, 2022/10/01/.
[56] A.S. Tote, S.S. Pardeshi, A.D. Patange, Automatic number plate detection using TensorFlow in Indian scenario: an optical character recognition approach, Mater. Today: Proc. (2022) https://doi.org/10.1016/j.matpr.2022.09.165, 2022/09/20/.
[57] D. Leykin, M. Lahad, L. Aharonson-Daniel, Gauging urban resilience from social media, Int. J. Disaster Risk Reduc. 31 (2018) 393–402 https://doi.org/10.1016/j.ijdrr.2018.04.021, 2018/10/01/.
[58] S. Li, L. Tong, C. Zhai, Extraction and modelling application of evacuation movement characteristic parameters in real earthquake evacuation video based on deep learning, Int. J. Disaster Risk Reduc. 80 (2022), 103213 https://doi.org/10.1016/j.ijdrr.2022.103213, 2022/10/01/.
[59] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. 28 (2015).
[60] J. Huang, et al., Speed/accuracy trade-offs for modern convolutional object detectors, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7310–7311.
[61] H. Sun, B. Dhingra, M. Zaheer, K. Mazaitis, R. Salakhutdinov, W.W. Cohen, Open Domain Question Answering Using Early Fusion Of Knowledge Bases And Text, 2018 arXiv preprint arXiv:1809.00782.
[62] H. Gunes, M. Piccardi, Affect recognition from face and body: early fusion vs. late fusion, in: 2005 IEEE International Conference on Systems, Man and Cybernetics, vol. 4, IEEE, 2005, pp. 3437–3443.
[63] Z. Ahmad, R. Jindal, N.S. Mukuntha, A. Ekbal, P. Bhattachharyya, Multi-modality helps in crisis management: an attention-based deep learning approach of leveraging text for image classification, Expert Syst. Appl. 195 (2022) 116626.
[64] R. Hou, H. Chang, B. Ma, S. Shan, X. Chen, Cross attention network for few-shot classification, Adv. Neural Inf. Process. Syst. 32 (2019).
[65] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[66] Y. Lu, et al., Automated arrhythmia classification using depthwise separable convolutional neural network with focal loss, Biomed. Signal Process Control 69 (2021), 102843 https://doi.org/10.1016/j.bspc.2021.102843, 2021/08/01/.
[67] J. Mushava, M. Murray, A novel XGBoost extension for credit scoring class-imbalanced data combining a generalized extreme value link and a modified focal loss function, Expert Syst. Appl. 202 (2022), 117233 https://doi.org/10.1016/j.eswa.2022.117233, 2022/09/15/.
[68] A.S. Dina, A. Siddique, D. Manivannan, A deep learning approach for intrusion detection in Internet of Things using focal loss function, Internet Things (2023) 100699.
[69] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[70] H. Kamangir, E. Krell, W. Collins, S.A. King, P. Tissot, Importance of 3D convolution and physics on a deep learning coastal fog model, Environ. Model. Software 154 (2022/08/01/2022) 105424, https://doi.org/10.1016/j.envsoft.2022.105424.
[71] J. Duell, X. Fan, B. Burnett, G. Aarts, S.-M. Zhou, A comparison of explanations given by explainable artificial intelligence methods on analysing electronic health records, in: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), IEEE, 2021, pp. 1–4.
[72] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable ai: a review of machine learning interpretability methods, Entropy 23 (1) (2020) 18.
[73] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
[74] M.H. Zarei, Managing medical waste in humanitarian supply chains: lessons for healthcare services, Br. J. Healthc. Manag. 28 (11) (2022) 269–274.
[75] A. Thomas, A. Urquhart, Global Humanitarian Assistance Report 2020, Summary-Development Initiatives, Bristol, 2020.
[76] N. Altay, G. Kovács, Innovation in humanitarian supply chains: a systematic review, Innovation 3 (2018) 1–24.
[77] M.H. Zarei, R. Carrasco-Gallego, S. Ronchi, To greener pastures: an action research study on the environmental sustainability of humanitarian supply chains, Int. J. Oper. Prod. Manag. 39 (11) (2019) 1193–1225.
[78] S. Huang, et al., Language Is Not All You Need: Aligning Perception With Language Models, 2023 arXiv preprint arXiv:2302.14045.
[79] H. Zhang, X. Li, L. Bing, Video-Llama: An Instruction-Tuned Audio-Visual Language Model For Video Understanding, 2023 arXiv preprint arXiv:2306.02858.
[80] J.-B. Alayrac, et al., Flamingo: a visual language model for few-shot learning, Adv. Neural Inf. Process. Syst. 35 (2022) 23716–23736.
[81] M.A. Hassanien, V.K. Singh, D. Puig, M. Abdel-Nasser, Predicting breast tumor malignancy using deep ConvNeXt radiomics and quality-based score pooling in ultrasound sequences, Diagnostics 12 (5) (2022) 1053.
[82] J. Qi, M. Nguyen, W. Yan, Waste classification from digital images using ConvNeXt, in: Pacific-rim Symposium on Image and Video Technology, 2022.
[83] H. Zhang, C. Liu, J. Ho, Z. Zhang, Crack detection based on convnext and normalization, in: Journal of Physics: Conference Series, vol. 2289, IOP Publishing, 2022, 012022 1.
[84] L.-Y. Wang, H.-H. Hu, L. Wang, J.-Q. Qin, Privacy assurances and social sharing in social commerce: the mediating role of threat-coping appraisals, J. Retailing Consum. Serv. 67 (2022), 103028 https://doi.org/10.1016/j.jretconser.2022.103028, 2022/07/01/.
[85] S. Abirami, D.G.K.D. Prasanna Venkatesan, Deep learning and spark architecture based intelligent brain tumor MRI image severity classification, Biomed. Signal Process Control 76 (2022) 103644, https://doi.org/10.1016/j.bspc.2022.103644.
[86] G. Neelima, D.R. Chigurukota, B. Maram, B. Girirajan, Optimal DeepMRSeg based tumor segmentation with GAN for brain tumor classification, Biomed. Signal Process Control 74 (2022) 103537, https://doi.org/10.1016/j.bspc.2022.103537.
[87] N.-R. Zhou, T.-F. Zhang, X.-W. Xie, J.-Y. Wu, Hybrid quantum–classical generative adversarial networks for image generation via learning discrete distribution, Signal Process. Image Commun. (2022), 116891 https://doi.org/10.1016/j.image.2022.116891, 2022/11/01/.
[88] H. Woldesellasse, S. Tesfamariam, Data augmentation using conditional generative adversarial network (cGAN): application for prediction of corrosion pit depth and testing using neural network, J. Pipeline Sci. Eng. (2022), 100091 https://doi.org/10.1016/j.jpse.2022.100091, 2022/10/25/.
[89] D. Mahmudnia, M. Arashpour, Y. Bai, H. Feng, Drones and blockchain integration to manage forest fires in remote regions, Drones 6 (11) (2022) 331.
[90] J. Ayoub, X.J. Yang, F. Zhou, Combat COVID-19 infodemic using explainable natural language processing models, Inf. Process. Manag. 58 (4) (2021), 102569 https://doi.org/10.1016/j.ipm.2021.102569, 2021/07/01/.
[91] C. Bosco, V. Patti, S. Frenda, A.T. Cignarella, M. Paciello, F. D'Errico, Detecting racial stereotypes: an Italian social media corpus where psychology meets NLP, Inf. Process. Manag. 60 (1) (2023), 103118 https://doi.org/10.1016/j.ipm.2022.103118, 2023/01/01/.