Multimodal Fusion for Disaster Event Classification on Social Media: A Deep Federated Learning Approach

Ayoub El-Niss
Department of Computer Science,
Jordan University of Science and Technology
Irbid, Jordan
amelniss20@cit.just.edu.jo

Ahmad Alzu'bi*
Department of Computer Science,
Jordan University of Science and Technology
Irbid, Jordan
agalzubi@just.edu.jo

Abdelrahman Abuarqoub
Cardiff School of Technologies,
Cardiff Metropolitan University,
Cardiff, Wales, UK
AAbuarqoub@cardiffmet.ac.uk

ABSTRACT

This paper explores the intersection of federated learning and disaster identification using a curated dataset of captioned images sourced from social media. Leveraging a federated learning framework, our methodology involves iterative client updates, server-side aggregation, and comprehensive testing to enhance the global model's understanding of disaster-related multimedia content. The study incorporates deep embeddings extracted and encoded by BERT models with generic image features extracted by ResNet, which is followed by a late fusion strategy to formulate discriminating features from both textual and visual modalities. Through collaborative efforts among decentralized clients, the global model demonstrates improved accuracy and robustness in identifying and classifying diverse disaster- related scenarios. With an accuracy of 85.1% and F1-score of 85.2%, this multimodal deep federated model contributes to the evolving field of federated learning, highlighting the significance of adaptability, data privacy preservation, and iterative feature refinement in improving the performance of disaster event identification and analysis.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; *Artificial intelligence*; • **Applied computing** → Physical sciences and engineering.

KEYWORDS

Federated Learning, Multimodal, Deep neural networks, Disaster identification, Social media.

ACM Reference Format:

Ayoub El-Niss, Ahmad Alzu'bi, and Abdelrahman Abuarqoub. 2023. Multimodal Fusion for Disaster Event Classification on Social Media: A Deep Federated Learning Approach. In *Proceedings of 7th International Conference on Future Networks & Distributed Systems (ICFNDS'23). December 21-22, 2023, Dubai, UAE. (ICFNDS'23).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding Author

ICFNDS'23, December 21, 2023, Dubai, UAE

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In recent years, the ubiquity of social media platforms has fundamentally transformed the way information is published and consumed during critical events such as natural disasters and wars. Users across the globe share real-time updates, images, and narratives, offering an invaluable source of data for disaster identification and response. The sheer volume and speed of information shared on social media, however, present a a significant challenge in efficiently and accurately categorizing and responding to relevant content [9][16].

However, there are significant obstacles in the way of obtaining useful information from social media data. Existing disaster classification models struggle to produce fast and reliable findings due to the large volume of data as well as its inherent heterogeneity and noise [17]. Traditional methods frequently depend on more general categories, which makes it difficult to grasp the subtleties of particular disaster types, such as "loss of life," which are essential for efficient response operations [32][7]. The ability to accurately identify and categorize specific types of disasters within the broader context of an event is critical for efficient and effective response. Knowing whether a disaster has caused "loss of life" or "fire" allows for targeted resource assignment, prioritization of rescue efforts, and publicizing of relevant information to the affected population [25][18][14].

In order to extract discriminating features from social data, more innovative and effective approaches are needed. Nevertheless, the majority of researchers tend to rely on unimodal data, particularly focusing on text or image data, when analyzing disaster events. Recent findings indicate that systems utilizing multimodal data outperform those relying solely on unimodal data [18][4][34][13]. Multimodal machine learning endeavors to construct models capable of correlating information from various modalities, constituting a dynamic and increasingly vital multidisciplinary field.

This paper proposes a multimodal deep federated learning paradigm to overcome the limitations of existing methods by utilizing a combination of multimodal fusion and Federated Learning (FL).

Multimodal fusion aims to combine information from different modalities, such as text and images, to achieve better performance than solely relying on individual modalities. By merging the textual descriptions in tweets with the visual information present in accompanying images, this approach provides a more comprehensive understanding of the disaster event. This leads to an enhanced feature space for the model, ultimately contributing to improved classification accuracy and robustness [30]. Federated learning offers a valuable advantage in this context by enabling the model to be trained on decentralized data residing on mobile devices. This not only addresses concerns regarding data privacy and security but also enables the model to leverage vast amounts of data that would otherwise be inaccessible [31][12]. Additionally, FL facilitates scalability, allowing the model to be continuously updated and improved as more devices contribute data, making it adaptable to dynamic disaster scenarios.

The remaining part of this paper is organized as follows: Section 2 reviews the related works, Section 3 presents the proposed methodology and pipeline of multimodal federated learning model, Section 4 discusses the experimental results and Section 5 concludes this work.

2 RELATED WORK

Over the past few years, social media has emerged as an invaluable source of real-time information during disaster events. Users often share text, and images related to the disaster, providing valuable insights for current state awareness, damage assessment, and resource allocation. However, effectively utilizing this multimodal data for disaster event classification remains challenging.

Early work on disaster event classification focused primarily on text analysis. Traditional machine learning techniques such as Support Vector Machines (SVM) [28] and Naive Bayes [22] have been implemented to classify textual content into different disaster categories [26]. However, these methods often fail to capture the nuances of natural language and may be subjected to noise and irrelevant information.

Deep learning has shown significant promise in improving the accuracy of text-based disaster event classification. Deep Neural Networks (DNNs) have been successfully used to extract complex features from text data, leading to better performance compared to traditional approaches [18][6]. However, relying solely on text cannot be reliable in situations where visual information is crucial. For instance, images and videos can provide valuable context for confirming the event type or how severe it is [25].

Recent research has explored the potential of multimodal data fusion for disaster event classification. This involves combining text and visual information to obtain a more comprehensive and informative event representation. Early fusion approaches [5] concatenate text and image features at the input layer, allowing them to be learned jointly by a single neural network. This has been shown to improve classification accuracy compared to unimodal approaches.

Late fusion approaches [11] combine the predictions obtained from separate text and image models at the classification level. This allows for more flexibility in model design and training, even though it can be less effective than early fusion in capturing complex relationships between modalities of crisis events in social media [1].

FL technologies have surfaced as an encouraging approach for training deep learning models on distributed data without compromising user privacy. In the context of disaster event classification, FL enables the training of a global model across multiple edge devices (e.g., mobile phones) [32][31][19]. This distributed training approach can be particularly beneficial for collecting and analyzing data from geographically spread regions affected by disaster events. Moreover, FL can address the data shortage issues often encountered in disaster scenarios.

Several recent works have investigated the application of FL or multimodal data with social media for disaster event analysis [32][18][14][34][1][23][29][33]. While the proposed approaches demonstrate promising advancements in multimodal fusion for disaster event classification and analysis, several challenges remain on the horizon, demanding further exploration and advanced solutions. First, data heterogeneity. The vast and ever-growing landscape of social media data presents a formidable challenge due to its inherent noise and heterogeneity [20]. This diversity, encompassing various text formats, image styles, and varying levels of information quality, necessitates robust models capable of effectively extracting meaningful features and generalizing to unseen data.

Second, privacy concerns. In the light of increasing data sensitivity, collecting and sharing disaster-related information raises ethical and legal concerns surrounding user privacy. Addressing these concerns requires the development of privacy-preserving techniques that ensure user anonymity while enabling the effective utilization of data for disaster response efforts [8].

Third, model Interpretability. Despite their impressive performance, deep learning models can often be nontransparent, shrouding their decision-making processes in a veil of mystery. This lack of transparency can hinder trust and limit their practical applications, particularly in critical situations like disaster response [15]. Future research should prioritize developing interpretable models that shed light on their reasoning and enable informed decision-making based on their outputs.

Fourth, is scalability. The dynamic nature of disaster scenarios demands efficient and robust infrastructure to facilitate large-scale deployment of machine learning models. This necessitates investigating distributed architectures and resource-efficient algorithms that can operate seamlessly under resource constraints and dynamic network conditions [17][7]. Our proposed approach leverages the complementary information present in text and images while ensuring data privacy and scalability through a federated deep-learning approach with social media.

3 METHODOLOGY

Figure 1 shows the pipeline of the proposed model for disaster event classification with social media analysis. The proposed approach consists of three main components: feature extraction, multimodal data fusion, and deep federated learning process. We first prepare the dataset of disaster events that includes textual tweets with images. Then we design

a feature extraction model with BERT [27] for textual features on Twitter and ResNet [24] for visual image features. Finally, a federated learning environment is configured with effective client-server deep learning setups to learn the multimodal discriminating descriptors and to classify disaster events automatically. These pipeline phases are further elaborated in the next subsections.

3.1 Dataset Preparation

We use the MEDIC [2] a public available dataset that has a total of 5831 image-tweet pairs and was divided into two parts: 5247 samples in the training set, and 584 samples for the test set. It

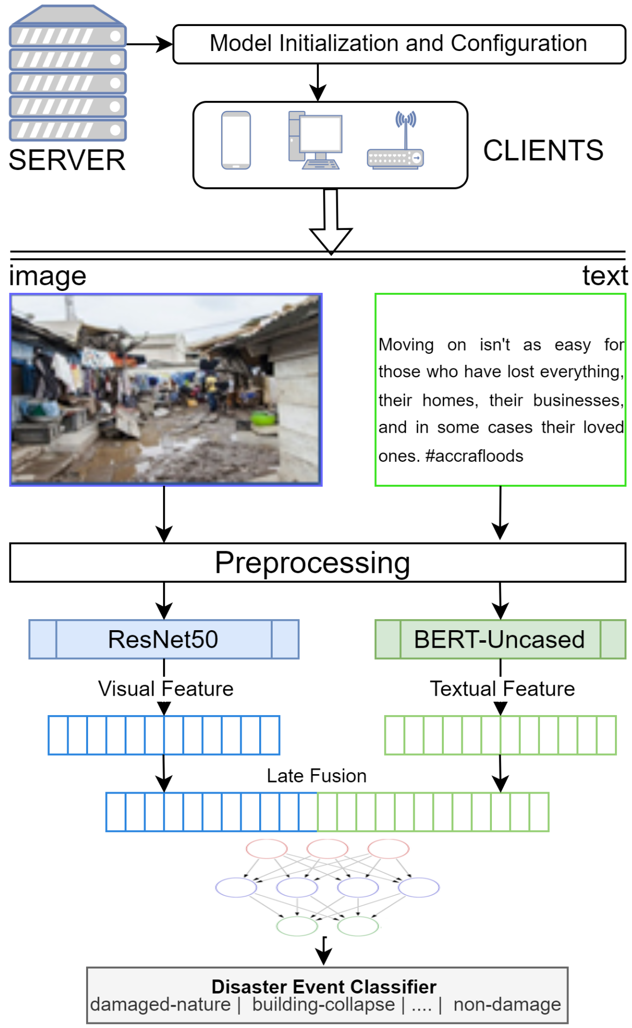

Figure 1: The proposed framework for disaster event classification with deep federated learning.

is used to evaluate models that can identify and classify different image-tweet pairs related to disasters. The essential property of this dataset is the imbalance of classes, which means that the distribution of samples between different classes is not equal. The highest number of samples is in the non-damage category, representing cases where no damage was detected, with 2975 cases. Figure 2 shows a sample disaster event represented by textual and image, i.e., multimodalities.

In contrast, the loss of life category, which refers to image-tweet pairs that show damage resulting in casualties during disasters, has the fewest samples, with only 240 cases. To facilitate the classification task, the dataset contains five distinct pairs of disaster images and tweets in addition to the non-damage category, the disaster Images and texts from the training dataset are presented to the proposed model batch by batch. For our work, the batch size is 12.

We conducted preprocessing on each image, resizing them to dimensions of 228x228x3. To ensure standardized pixel values, normalization techniques were applied, scaling the image pixels within the range of 0 to 1. This involved normalizing the pixel values using mean and standard deviation parameters of [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225], respectively. This process standardizes the pixel values, establishing a mean of 0 and a standard deviation of 1, which aids in stabilizing the training process. Additionally, various augmentation techniques were employed to enhance the dataset, including 1) random horizontal flip to present the model with varied perspectives on the image objects, 2) color jitter that introduces random variations to the color channels of disaster-related images to bolster the model's adaptability, and 3) Random rotation to provide a variety of the orientation of the disaster event images.

3.2 Multimodal Feature Extraction

For the visual features, we apply a transfer learning procedure by implementing a ResNet50 pre-trained model. ResNet50 is a deep convolutional neural network (CNN) architecture and a powerful image classification model that can be trained on large datasets and achieve state-of-the-art results. One of its key innovations is the use of residual connections, which allow the network to learn a set of residual functions that map the input to the desired output.

Figure 2: Sample tweet demonstrating image, caption, and tags (labels).

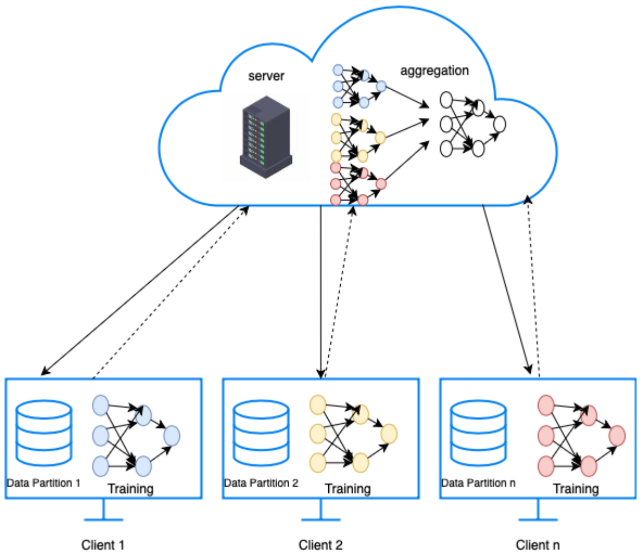

Figure 3: The generic setup of the federated learning on each round.

These residual connections enable the network to learn much deeper architectures than was previously possible, without suffering from the problem of vanishing gradients. The architecture of ResNet50 is divided into four main parts: the convolutional layers, the identity block, the convolutional block, and the fully connected layers. The convolutional layers are responsible for extracting features from the input image, while the identity block and convolutional block are responsible for processing and transforming these features. Finally, the fully connected layers are used to make the final classification.

For the textual features, we extract meaningful features from the textual content of Tweets in the dataset using BERT. BERT is a leading natural language processing (NLP) model that transforms words into numerical representations, fundamentally altering the landscape of language understanding. Operating on a transformer-based [22] architecture, BERT introduces a bidirectional approach to analyze the context of words in a sentence. BERT's strength lies in its ability to concurrently consider both preceding and succeeding words, capturing nuanced semantic relationships within the text. The model undergoes pre-training on extensive corpora to acquire contextual representations, followed by fine-tuning for our specific classification task, facilitating accurate predictions and enhancing its adaptability to diverse textual tweets.

In our experiments, we employ the BERT-base-uncased variant [3], a model renowned for its efficacy in handling diverse textual data. This variant enables the training of machine-learning models on Twitter data, offering a robust foundation for feature extraction and subsequent analysis by utilizing BERT, we harness the power of contextual embeddings to enhance the understanding of Tweets, enabling our models to capture intricate linguistic nuances and contribute to more accurate disaster- related content identification.

3.3 Data Fusion and Classification

The outcomes generated by the dense layer in both the visual and textual components are amalgamated to establish a unified representation that encapsulates both visual and text features. Employing a late fusion approach [10], we concatenate visual and textual elements, enabling the attainment of a comprehensive representation at a deep level. Late fusion involves the initial independent analysis of inputs from multiple modalities, followed by the integration of outputs or features from these modalities to arrive at a conclusive prediction. In the ultimate classification layer of both modalities, an identical number of hidden nodes, specifically 512 nodes, is utilized. This uniform size ensures equitable contributions from both the textual and visual aspects.

3.4 Federated Learning Setup

The federated learning process begins with the initialization of a global model which is our multimodal model and individual client models, as shown in Figure 3. Key hyperparameters, such as the number of training rounds, the selected clients per round, and the training epochs are set. This establishes the foundation for the collaborative learning process across decentralized datasets. The global model represents the collective knowledge to be enhanced through the iterative collaboration of individual client models.

In each iteration of the federated learning loop, a subset of clients is randomly chosen to participate in the model update. This process simulates the decentralized nature of FL, reflecting real- world scenarios where clients autonomously process and update their models locally. The client update function is executed for each selected client, ensuring that each participant contributes to the overall improvement of the global model. Subsequently, the updated client models are aggregated at the server using a mean aggregation method [21]. This collaborative effort harmonizes the individual insights gained by each client, fostering a shared understanding embedded in the refined global model. The global model's performance is then evaluated on a separate test subset, providing insights into its generalization capabilities and overall effectiveness.

Throughout the federated learning process, training and testing metrics are meticulously recorded and analyzed. These metrics include training loss, test loss, accuracy, precision, recall, and F1 score. Continuous monitoring allows for the observation of convergence patterns and the overall progress of the federated learning model. By examining these metrics over multiple rounds, insights into the model's learning dynamics and collaborative efficiency emerge. The final conclusions drawn from the federated learning process consider the achieved performance, the convergence trends, and the collaborative knowledge integration among the decentralized clients.

4 EXPERIMENTS AND RESULTS

In this section, we first describe the implementation configuration of the proposed method, we then show the performance of our model, followed by results discussion.

4.1 Implementation Configuration

Our approach was subjected to a comprehensive evaluation across a training and testing split by 80% and 20% respectively and spanning over 31 rounds. The experiments were executed on a high-performance computational cluster equipped with A100 GPUs, enabling the seamless training of complex neural network models. To achieve this, we leveraged the computational power of PyTorch, a prominent deep learning framework, ensuring seamless integration with state of the art models.

Each training round involved the configuration of the model, where a random selection of 6 clients out of 20 possible clients participated with 10 epochs for each client. We used the Stochastic Gradient Descent (SGD) optimizer with learning rate set to 0.1. The deep federated learning setup employed a robust mechanism for aggregating client models. The server aggregation function calculated the mean of the weights of client models, updating the global model and subsequently distributing the aggregated global model to each participating client.

The performance was measured by testing the global model on the test set in each round. The training dataset, a vital element in our training process, underwent a shuffling process and was subsequently distributed among the participating clients. The shuffling process ensures a diverse representation of the data, contributing to the robustness and generalizability of our proposed approach.

4.2 Model Performance

Table 1 shows the performance of the proposed model on the test set averaged over a range of 31 rounds, highlighting the scores obtained in every 3 rounds. The convergence and refinement of the model are evident in the evolution of key metrics across rounds.

In the initial round, the model exhibited promising capabilities with an accuracy of 79.6%, precision of 81.7%, recall of 79.6%, and an F1 score of 80.3%. As the federated learning process advanced, our model consistently demonstrated improved performance. Notably, by the 10^{th} round, the accuracy surged to 84.4%, with precision, recall, and F1 score reaching 85.6%, 84.4%, and 84.7%, respectively. This trend continued, with the model accuracy peaked at 85.1% in the 30^{th} round, asserting its capability in multimodal classification. best results on the 30^{th} round, and the result of the 24^{th}-29^{th} rounds is slightly less. This indicates that the model start showing a stable generalization capability on the training data with a slight change in the classification metrics. In this sense, if the 9^{th} round, for example, introduced new or diverse data, it could have contributed to improved model performance. Likewise, if the 12^{th} round encountered challenging or unrepresentative instances, the model might have struggled to generalize.

The consistent upward trend in critical performance metrics, such as test accuracy, precision, recall, and F1 score, observed throughout the experimental trials. This progression suggests a continuous learning and generalization process, demonstrating the model's capacity to refine its predictive capabilities over multiple training rounds.

One particularly powerful trend is the gradual improvement in test accuracy from an initial 79.6% in the 0^{th} round to a notable peak of 85.1% in the 31^{th} round, as visually depicted in Figure 1. This substantial increase indicates the model's ability to learn

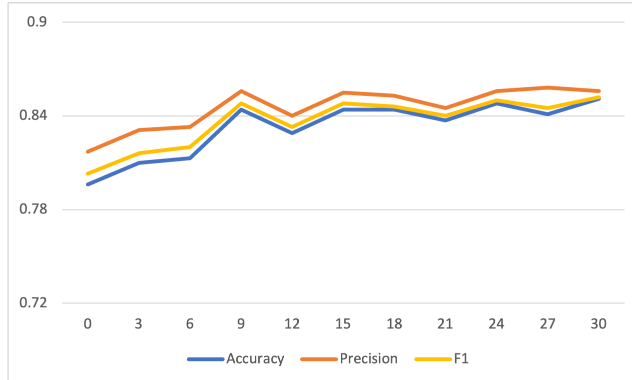

Figure 4: Model Performance Evolution. A demonstration of a consistent improvement in test accuracy, precision, and F1 score.

from the federated dataset and make more accurate predictions over time. The simultaneous elevation in precision, recall, and F1 score, as illustrated in the graph, further reinforces the model's efficacy in balancing between correctly identifying disaster events and minimizing false positives.

Table 1: The performance of the proposed model on different rounds.

Round	Accuracy	Precision	Recall	F1
0	0.796	0.817	0.796	0.803
3	0.810	0.831	0.810	0.816
6	0.813	0.833	0.813	0.820
9	0.844	0.856	0.844	0.848
12	0.829	0.840	0.829	0.833
15	0.844	0.855	0.844	0.848
18	0.844	0.853	0.844	0.846
21	0.837	0.845	0.837	0.840
24	0.848	0.856	0.848	0.850
27	0.841	**0.858**	0.841	0.845
30	**0.851**	0.856	**0.851**	**0.852**

5 CONCLUSION

In conclusion, this paper presents a multimodal deep federated learning model to address the challenges of categorizing and responding to critical events on social media. The approach integrates textual descriptions with visual data to enhance the model's understanding of disaster events. Leveraging federated learning, the model is trained on decentralized data, ensuring privacy and scalability. The experimental results demonstrate the effectiveness of our approach, showing consistent improvements in critical performance metrics. The stability in several classification scores highlights the robustness of the proposed method. This approach holds promise for advancing real-world problems by overcoming existing limitations in data availability and privacy concerns and leveraging the multimodal fusion and federated learning.

REFERENCES

[1] Mahdi Abavisani, Liwei Wu, Shengli Hu, Joel Tetreault, and Alejandro Jaimes. 2020. Multimodal categorization of crisis events in social media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14679–14689.

[2] Firoj Alam, Tanvirul Alam, Md Arid Hasan, Abul Hasnat, Muhammad Imran, and Ferda Ofli. 2023. MEDIC: a multi-task learning dataset for disaster image classification. *Neural Computing and Applications* 35, 3 (2023), 2609–2632.

[3] Firoj Alam, Tanvirul Alam, Md Arid Hasan, Abul Hasnat, Muhammad Imran, and Ferda Ofli. 2023. MEDIC: a multi-task learning dataset for disaster image classification. *Neural Computing and Applications* 35, 3 (2023), 2609–2632.

[4] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* 41, 2 (2018), 423–443.

[5] George Barnum, Sabera Talukder, and Yisong Yue. 2020. On the benefits of early fusion in multimodal representation learning. *arXiv preprint arXiv:2011.07191* (2020).

[6] Andrew Carlin, Mohammad Hammoudeh, and Omar Aldabbas. 2015. Intrusion detection and countermeasure of virtual cloud systems-state of the art and current challenges. *International Journal of Advanced Computer Science and Applications* 6, 6 (2015).

[7] Neha Chaudhuri and Indranil Bose. 2020. Exploring the role of deep neural networks for post-disaster decision support. *Decision Support Systems* 130 (2020), 113234.

[8] Kejun Chen, Shuai Zhang, Zhikun Li, Yi Zhang, Qingxu Deng, Sandip Ray, and Yier Jin. 2018. Internet-of-Things security and vulnerabilities: Taxonomy, challenges, and practice. *Journal of Hardware and Systems Security* 2 (2018), 97–110.

[9] Lokabhiram Dwarakanath, Amirrudin Kamsin, Rasheed Abubakar Rasheed, Anitha Anandhan, and Liyana Shuib. 2021. Automated machine learning approaches for emergency response and coordination via social media in the aftermath of a disaster: A review. *IEEE Access* 9 (2021), 68917–68931.

[10] Konrad Gadzicki, Razieh Khamsehashari, and Christoph Zetzsche. 2020. Early vs late fusion in multimodal convolutional neural networks. In *2020 IEEE 23rd international conference on information fusion (FUSION)*. IEEE, 1–6.

[11] Nathan Gaw, Safoora Yousefi, and Mostafa Reisi Gahrooei. 2022. Multimodal data fusion for systems improvement: A review. *IISE Transactions* 54, 11 (2022), 1098–1116.

[12] Mohammad Hammoudeh, Gregory Epiphaniou, Sana Belguith, Devrim Unal, Bamidele Adebisi, Thar Baker, ASM Kayes, and Paul Watters. 2020. A service-oriented approach for sensing in the Internet of Things: Intelligent transportation systems and privacy use cases. *IEEE Sensors Journal* 21, 14 (2020), 15753–15761.

[13] Mohammad Hammoudeh, Robert Newman, Christopher Dennett, Sarah Mount, and Omar Aldabbas. 2015. Map as a service: a framework for visualising and maximising information return from multi-modal wireless sensor networks. *Sensors* 15, 9 (2015), 22970–23003.

[14] Md Yasin Kabir and Sanjay Madria. 2019. A deep learning approach for tweet classification and rescue scheduling for effective disaster management. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 269–278.

[15] Abhinav Kumar, Jyoti Prakash Singh, and Sunil Saumya. 2019. A comparative analysis of machine learning techniques for disaster-related tweet classification. In *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129)*. IEEE, 222–227.

[16] Vasileios Linardos, Maria Drakaki, Panagiotis Tzionas, and Yannis L Karnavas. 2022. Machine learning in disaster management: recent developments in methods and applications. *Machine Learning and Knowledge Extraction* 4, 2 (2022).

[17] Sally Lu, Gordon A Christie, Thanh T Nguyen, Jeffrey D Freeman, and Edbert B Hsu. 2022. Applications of artificial intelligence and machine learning in disasters and public health emergencies. *Disaster medicine and public health preparedness* 16, 4 (2022), 1674–1681.

[18] Ferda Ofli, Firoj Alam, and Muhammad Imran. 2020. Analysis of social media data using multimodal deep learning for disaster response. *arXiv preprint arXiv:2004.11838* (2020).

[19] Sharnil Pandya, Gautam Srivastava, Rutvij Jhaveri, M Rajasekhara Babu, Sweta Bhattacharya, Praveen Kumar Reddy Maddikunta, Spyridon Mastorakis, Md Jalil Piran, and Thippa Reddy Gadekallu. 2023. Federated learning for smart cities: A comprehensive survey. *Sustainable Energy Technologies and Assessments* 55 (2023), 102987.

[20] Viktor Pekar, Jane Binner, Hossein Najafi, Chris Hale, and Vincent Schmidt. 2020. Early detection of heterogeneous disaster events using social media. *Journal of the Association for Information Science and Technology* 71, 1 (2020), 43–54.

[21] Krishna Pillutla, Sham M Kakade, and Zaid Harchaoui. 2022. Robust aggregation for federated learning. *IEEE Transactions on Signal Processing* 70 (2022), 1142–1154.

[22] Mucahid Mustafa Saritas and Ali Yasar. 2019. Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International journal of intelligent systems and applications in engineering* 7, 2 (2019), 88–91.

[23] Y Supriya and Thippa Reddy Gadekallu. 2023. Particle Swarm-Based Federated Learning Approach for Early Detection of Forest Fires. *Sustainability* 15, 2 (2023), 964.

[24] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.

[25] Haiman Tian, Hector Cen Zheng, and Shu-Ching Chen. 2018. Sequential deep learning for disaster-related video classification. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 106–111.

[26] Manabu Torii, Lanlan Yin, Thang Nguyen, Chand T Mazumdar, Hongfang Liu, David M Hartley, and Noele P Nelson. 2011. An exploratory study of a text classification framework for Internet-based surveillance of emerging epidemics. *International journal of medical informatics* 80, 1 (2011), 56–66.

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[28] Lipo Wang. 2005. *Support vector machines: theory and applications*. Vol. 177. Springer Science & Business Media.

[29] Zheye Wang and Xinyue Ye. 2018. Social media analytics for natural disaster management. *International Journal of Geographical Information Science* 32, 1 (2018), 49–72.

[30] Yao-Yuan Yang, Cyrus Rashtchian, Hongyang Zhang, Russ R Salakhutdinov, and Kamalika Chaudhuri. 2020. A closer look at accuracy vs. robustness. *Advances in neural information processing systems* 33 (2020), 8588–8601.

[31] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. 2021. A survey on federated learning. *Knowledge-Based Systems* 216 (2021), 106775.

[32] Zehui Zhang, Ningxin He, Dongyu Li, Hang Gao, Tiegang Gao, and Chuan Zhou. 2022. Federated transfer learning for disaster classification in social computing networks. *Journal of safety science and resilience* 3, 1 (2022), 15–23.

[33] Zehui Zhang, Ningxin He, Dongyu Li, Hang Gao, Tiegang Gao, and Chuan Zhou. 2022. Federated transfer learning for disaster classification in social computing networks. *Journal of safety science and resilience* 3, 1 (2022), 15–23.

[34] Zhiqiang Zou, Hongyu Gan, Qunying Huang, Tianhui Cai, and Kai Cao. 2021. Disaster image classification by fusing multimodal social media data. *ISPRS International Journal of Geo-Information* 10, 10 (2021), 636.