Selection of Dataset for Emotion Detection with Respect to Federated Learning

G. K. Jakir Hussain[1] and G. Manoj[2]

[1]*Division of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Karunya Nagar, Coimbatore 641114, Tamil Nadu, India and Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology, Coimbatore 641407, Tamil Nadu, India*

[2]*Division of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Karunya Nagar, Coimbatore 641114, Tamil Nadu, India*

Keywords: Emotion Detection, Federated Learning, Dataset, Data Diversity, Human Computer Interactions.

Abstract: Emotion detection (ED) plays a vital role in applications like health, human–computer interaction (HCI), and personalization. Federated learning (FL) has the potential to provide robust models for ED while keeping data private across distributed clients. This work has considered the critical factors that influence dataset selection for FL on ED tasks. The range of dataset types, class balance for varying emotional states, and attention to privacy considerations that safeguard users' sensitive information are among the important factors to consider. Therefore, several datasets are examined in terms of how well they extent the range of emotional expressions and replicate actual client data distributions. This study highlights datasets such as FER-2013 for photos, RAVDESS for audio, and ISEAR for textual data are highly relevant for the construction of generalized ED models in FL setups that simulate near-real-world settings, as per review publications. The FL technique can balance the data diversity and privacy preservation and hence saves a pathway to harness collective intelligence from distributed data sources while ensuring ethical practices for handling data. Finally, FL is becoming a critical topic for addressing issues in identifying ED by selecting the suitable datasets.

1 INTRODUCTION

Emotion detection (ED) is a process that identifies and classifies human emotions into modalities such as text, audio, and images. This requires methods for text by way of natural language processing (NLP) methods and those for audio through speech analysis, facial expression recognition (FER) from images, and more advanced ones that use deep learning (DL) and machine learning (ML) algorithms. Specifically, in ED through federated learning (FL), models are trained across decentralized devices, ensuring user privacy since data remains local while model updates are aggregated at a central location. It helps in improving mental health support, enhancing human computer interaction, and making the marketing strategies more personalized. Facilitating early identification of emotional disorders, it makes Artificial Intelligence (AI) empathetic and personalizes marketing campaigns according to the user's emotions for better engagement and customer satisfaction.

FL enables the training of decentralized models across a number of devices while maintaining local privacy of data, and at the same time, it centrally aggregates updates to models for better performance. FL is an approach to ML whereby model training takes place across decentralized devices holding local data samples, without transferring data to a central server (G. K. J. Hussain and G. Manoj, 2022). Key principles include data privacy, where data resides on local devices; iterative model updates, where only model parameters or gradients are shared and aggregated to create a global model; and security features enhancing scalability. The interaction between ED and FL gives privacy-preserving models that can analyze emotional states from decentralized devices, hence providing personalized experience to users with the safety of sensitive data. FL applied to ED faces several obstacles, including heterogeneity in device data, ineffective communication, and bias against the model due to unequal data distributions. Selecting a dataset for ED in FL involves privacy, diversity, and representativeness. The dataset needs

Hussain, G. K. J. and Manoj, G.
Selection of Dataset for Emotion Detection with Respect to Federated Learning.
DOI: 10.5220/0013917400004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 593-599
ISBN: 978-989-758-777-1

text, speech, and facial data for emotional cues. Anonymized data with consent is crucial for privacy. Diverse and representative datasets aid model performance across demographics. Public datasets like IEMOCAP and Sentiment140 may be adapted for FL. Data partitioning is essential for non-independent identity (non-iid) data in FL to ensure effective generalization.

2 EMOTION DETECTION

ED involves the identification and subsequent classification of human feel through various inputs, either text, speech, or facial expression. The techniques from NLP, Computer Vision, and ML are attached in analysis of emotional signals. ED can be categorized into three main types. ED can express feelings like joy or anger through text, speech, and facial recognition. Text-based detection uses NLP for sentiment analysis, while speech-based methods analyze tone and pitch in audio. Facial expression-based detection employs computer vision to identify emotions through facial movements. These approaches can be used combined or separately to enhance accuracy of ED systems in applications like customer service and mental health monitoring.

2.1 Types of Datasets

Type, diversity, and balance of emotional label are important considerations when choosing datasets for FL based on ED. For a strong model training, one would desire datasets that can be expressive about various journeys for emotional expression. Privacy is of the utmost importance, and datasets must adhere to regulations controlling the suppression of sensitive data. A dataset that can enable realistic client data distributions is also necessary in order to simulate real-world scenarios in FL. FER-2013 is used for images, RAVDESS for audio, and ISEAR for textual data. All of these offers decentralised learning and are helpful in acquiring some understanding of the emotional state.

2.1.1 Text Based ED

Selecting a dataset for text-based ED should be done in accordance with its relevance, balance, and diversity. It will involve a combination of labeled and emotive content, for example, for social media posts or reviews, or primed conversations. These range from widely used and richly emotional-annotation ones like ISEAR (International Survey on Emotion Antecedents and Reactions) and EmoReact. The real-world scenarios should be represented in the dataset for the model to build resilience. Moreover, datasets would be split to simulate scenarios of real-world FL enabling decentralized training of a model preserving confidential and private information associated with several users.

2.1.2 ISEAR Dataset

The ISEAR dataset is a unique database of 7,666 emotional statements contributed by 1,096 participants belonging to different cultural backgrounds. The dataset can be used in creating an Emotion Dominant Meaning Tree for the classification of emotional statements to obtain higher precision with increased emotion categories. The algorithms will be trained to learn how to identify and classify expressed emotions using the ISEAR Dataset from statements provided by participants. About 60% of the dataset is used in constructing the Emotion Dominant Meaning Tree. Jain and Asawa (Jain, S. and Asawa, K, 2019). mention the following strengths of the ISEAR dataset: it holds a large number of reports from respondents who expressed a wide range of emotions within a variety of settings. A deep analysis of emotional states is possible to be held due to the links between each emotion and certain event assessments and responses. It subsequently assures the model's effectiveness by selecting relevant records to increase the elicitation conditions' accuracy, thus proving to be a valuable dataset in case scenarios of real-world tests and validation of elicitation rules.

Table 1: Emotions described in ISEAR dataset. (Source: author).

Types of Emotions	Number of examples
Anger	1096
Disgust	1096
Fear	1095
Sadness	1096
Shame	1096
Joy	1094
Guilt	1093
Total	7666

This is based on the ISEAR dataset, which Alotaibi (Alotaibi, F.M., 2019) used to recognize emotions from text with logistic regression, which has since gone on to be categorically distinguished between fear, joy, sadness, guilt, and humiliation. Assessment metrics include the F1-score, precision, and recall. Better performance can be seen with logistic regression compared to the rest: support vector classifier, K-nearest neighbor, and extreme gradient boosting. Table 1 displays the emotions described in ISEAR dataset.

Asghar et al. (2019): The primary aim is the internet information identification of emotions through the effective utilization of supervised ML algorithms. It considers, in an assessment of the ML classifier, ISEAR data that contains 5477 reviews, basically categorized as joy, fear, sadness, humiliation, and guilt; this is after applying various preprocessing techniques on them, such as tokenization and stop word removal, before applying the classifier. Table 2 shows the detail description of ISEAR dataset and figure 1 depicts the types of anger based on the graphical representation of emotions from ISEAR dataset.

Table 2: Detail description of the ISEAR dataset. (Source: author).

Citation	Variety	Explanation	Real-world Pertinency
[2]	Various settings for eliciting emotions are modeled	Assumed to be high because of the construction of the cognitive-emotive	Relevant to cognitive systems that simulate emotions
[3]	Moderate, encompassing fundamental feelings	Manual annotation with distinct emotional descriptors	Helpful for classifying emotions using logistic regression
[4]	High, based on many web sources	High, utilising methods for supervised ML	Appropriate for identifying feelings in a variety of web material
[5]	Various conditions for eliciting emotions are modelled	High, making use of BERT's aptitudes	Efficient at recognising emotions from text

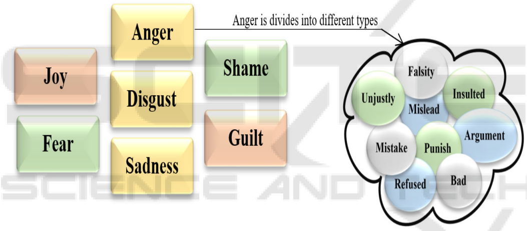

Figure 1: Graphical representation of the emotions in ISEAR dataset.

Among the many uses for the dataset in Adoma et al. 2020 is the assessment of machine learning models for emotion categorization. This dataset is ISEAR, consisting of 7666 sentences labeled for seven emotions. By making the dataset balanced, training and testing are possible without mitigation for class imbalance. As such, this dataset will have several preprocessing tasks such as tokenization or stop word removal. Results from this study may be compared to results from other studies in the same field, particularly serving as a benchmark for future ED studies with the application of ML models.

2.1.3 Speech Based ED

The audio-visual data in IEMOCAP dataset consists of 12 hours, which contains data from 10 actors in 5 sessions. It contains data that was both improvised and scripted, and it is divided into four emotional classes: neutral, happy, sad, and angry. Three to four assessors comment each utterance, and they award labels by majority vote. This multimodal dataset makes use of text, audio, and motion capture (Mocap) information. Its vast nature and good quality make it extremely important to research on ED. Using the IEMOCAP dataset, research by Tripathi and Beigi focuses on multi-modal emotion identification. This dataset allows neural network models for robust emotion recognition to be trained and evaluated. It consists of a variety of audio, video, and text data from dyadic interactions. The 1440 audio-only WAV files in the RAVDESS Emotional Speech Audio dataset have a sampling rate of 16 bits @ 48 kHz. Recorded in a neutral North American accent, it contains 24 professional actors 12 female, 12 male delivering lexically-matched phrases. Emotions manifest at two intensity levels (normal and strong), with a neutral expression in between. Calm, pleased, sad, furious, afraid, surprised, and disgusting are among the emotions. The RAVDESS dataset has been widely used in recent research to improve voice emotion recognition techniques. Table 3 displays the Comparison of the four datasets based on Speech based ED type.

Table 3: Comparison of the Four Datasets Based on Speech Based Ed (Source: Author).

Dataset	IEMOCAP	RAVDESS	CREMA-D	EmoDB
Explanation	Emotion speech database	Emotional speech audio	Emotional speech	Emotional speech
Source	USC Institute for Creative Technologies.	Ryerson Audio-Visual Database of Emotional Speech	Crowd-sourced	Technische Universität Berlin
Contributors	10 actors	24 actors	91 actors	10 actors
Reactions	Anger, Happiness, Sadness, Neutral, Surprise, Frustration, Excitedness, and Fear.	Calm, Happy, Sad, Angry, Neutral, Fearful, Disgust, Surprised.	Angry, Disgusted, Fear, Happy, Neutral, Sad, Surprise	Anger, Boredom, Disgust, Anxiety/Fear, Happiness, Sadness

A varied dataset called CREMA-D has 7,442 original audio clips with 91 performers representing a range of demographics (48 men and 43 women, ages 20 to 74, and different races). Twelve standardized words representing six different emotions Anger, Disgust, Fear, Happy, Neutral, Sad across four intensity ranges Low, Medium, High, and Unspecified were uttered by each actor. This dataset has a broad demographic representation and is used for voice emotion recognition research.

An innumerable of research into emotion recognition depends on the CREMA-D dataset. Shahzad et al. utilize CREMA-D for multi-modal deep learning, where text and audio inputs are fused to capture complex emotional expressions. The variation of this dataset substantially enhances model generalization across different modalities and thereby improves both the accuracy and the system's robustness in recognizing emotions.

The EMODB database developed by the Institute of Communication Science at the Technical University of Berlin which is an open-source German emotional database. It contains 535 utterances from ten professional speakers five of whom are male and the other five are female. The seven emotions covered by this database are: neutral, disgust, happiness, sorrow, anxiety, anger, and boredom. The data was down-sampled to 16 kHz for standardization after being initially captured at a 48 kHz sampling rate.

2.1.4 Facial Recognition-Based ED

Another The process of facial recognition-based ED trains algorithms to recognize human emotions from facial expressions using datasets like the Extended Cohn-Kanade Dataset (CK+) , the Facial Expression Recognition 2013 (FER2013), the Multimodal Database of Emotionally Salient Stimuli (MMI), the Japanese Female Facial Expression (JAFFE) Dataset, and the Affect Net Dataset. Labeled datasets for posed and naturalistic expressions are provided by CK+ and FER2013. MMI offers audio-visual signals for analyzing emotions, and JAFFE offers insights through the expressions of Japanese women.

The CK+ dataset includes 920 photos that have been modified and taken from the original CK+ dataset. These pictures have been pre-processed using the `haarcascade_frontalface_default' method to provide standard 48x48 pixel grayscale dimensions and face cropping. A Haar classifier was used to filter noisy images that were impacted by things like skin tone fluctuations, hair artifacts, and room illumination in order to improve clarity and identification. The dataset is organized into three columns for each entry: pixel values (2304 per image), emotion label (which has predefined indices for various emotions), and usage classification into three categories: training (80%), public test (10%), and private test (10%). Table 4 shows the Demonstration of the CK+ dataset and the number of samples.

Table 4: Demonstration of the CK+ dataset and the number of samples (Source: Author).

Types of Emotions in CK+ dataset	Number of Samples
Anger	45
Disgust	59
Fear	25
Happiness	69
Sadness	28
Surprise	83
Neutral	593
Contempt	18

The FER2013 dataset, which includes approximately 35,000 grayscale photos classified into seven emotional categories anger, disgust, fear, pleasure, sorrow, surprise, and neutral is used for FER. It was chosen specifically for use in FER task training and evaluation models. Every image of 48x48 pixels is labeled with different emotion categories. Researchers frequently utilize FER2013 to benchmark algorithms for applications involving computer vision and emotion identification. A number of DL-based facial ED studies are thus heavily reliant on the FER-2013 dataset. By incorporating audio, visual, and textual modalities, the Multimodal Emotion Lines Dataset (MELD) builds upon the Emotion Lines dataset. It features numerous speakers and more than 1,400 lines and 13,000 utterances taken from the Friends television series. Every word is annotated both with sentiment positive, negative, or neutral along with an emotion: Anger, Disgust, Sadness, Joy, Neutral, Surprise, or Fear. The dataset includes 213 TIFF-formatted, high-resolution grayscale pictures of 10 Japanese female expressers wearing 7 posed facial expressions, including neutral. Every expression has different photos from each expresser, for a total of many images per emotion. It contains 60 Japanese viewers averaged semantic judgments of six different face expressions. Table 5 shows the Comparison of the five datasets based on the Facial recognition-based ED.

Table 5: Comparison of the five datasets based on the facial recognition-based ED (Source: Author).

Dataset	CK+ [10]	FER2013	MMI	JAFFE	AffectNet
Source	Kaggle	Kaggle	Kaggle	Kaggle	Kaggle
Determination	FER	FER	ED	FER	FER
Quantity of Images	593	35,887	18,000 (video frames)	213	1,000,000+
Image Category	Grayscale, posed	Grayscale, posed	Video frames (multimodal)	Grayscale, posed	RGB, diverse
Resolution	Various	48x48 pixels	Various	256x256 pixels	Various
Reaction types	8 (fear, anger, disgust, surprise, contempt, happiness, neutral, sadness)	7 (fear, angry, surprise, disgust, happy, neutral, sad)	7 (anger, fear, surprise, disgust, sadness, neutral, happiness)	7 (disgust, anger, sadness, happiness, neutral, surprise, fear)	8 (fear, neutral, surprise, happy, anger, sad, contempt, disgust)
Notation	FACS action units	Crowd-sourced labels	Video-based labels	Human labeled	Crowd-sourced labels

2.1.5 Highly Used Datasets in ED

A wide range of datasets specialized to various modalities, including text, speech, and video, are extremely beneficial to emotion recognition research. Every dataset has a unique function: SemEval-2018 task 1 employs labelled tweets for text-based emotion analysis, whereas IEMOCAP offers emotional annotations in voice scenarios. For tasks involving the recognition of emotions in videos, Emo React is perfect because it comes with annotated clips that cover a wide range of emotions. MELD facilitates study into multimodal emotion comprehension by integrating emotion labels with textual dialogues from the Friends TV series. Emo Bank provides dimensional emotion rankings that are taken from blog sentences and social media, which gives a distinctive viewpoint. The versatility of SEMAINE's records, which allow for the examination of emotional states in text, video, and audio modes, is its main strength. These varied datasets are essential for creating and assessing ED models, which advances NLP and affective computing. Federated learning is the most common approach for emotion detection because it maintains decentralized data, improves model robustness by utilizing a variety of data sources, and facilitates collaboration among dispersed clients without jeo pardising sensitive data. The figure 2 shows some challenges in ED Dataset.

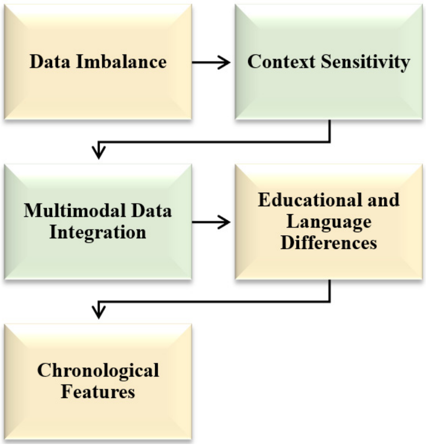
Data Imbalance
Context Sensitivity
Multimodal Data Integration
Educational and Language Differences
Chronological Features

Figure 2: Challenges in Ed Dataset.

2.2 Federated Learning

FL is a decentralized ML technique that uses local input and allows several devices or edge servers to cooperatively train a common model. Local computation and aggregates are used to compute model changes rather than transferring data to a central server. This lowers communication costs while protecting user privacy. FL comes extremely handy in situations like healthcare, IoT, and mobile applications when there are several, diverse datasets dispersed over various places. It solves problems like connectivity problems and data privacy laws, making it appropriate for applications needing customised models without sacrificing the confidentiality of individual data.

2.2.1 Federated Learning: Dataset Necessities

In FL, datasets must meet requirements for effective and secure decentralized model training by ensuring privacy and security through encryption and secure aggregation techniques. Differential privacy adds noise to data or model updates to protect against re-identification. Following FL protocols like Federated Averaging (FedAvg) shares only model parameters, not raw data, during updates. Handling data heterogeneity in FL involves addressing non-independent and identically distributed data with personalized models, data clustering, or federated transfer learning. Robust aggregation techniques like Trimmed Mean, Median or Krum help manage outliers or skewed data distributions. To enhance model performance, adaptive learning rates make adjustments in response to data heterogeneity. Protocols adapting to network conditions reduce communication overhead and transmission frequency. Scalability in FL supports a large number of devices by distributing workload, using hierarchical FL with local coordinators, and employing efficient resource management and load balancing techniques. Addressing these requirements enables effective leveraging of decentralized datasets while maintaining privacy, handling data heterogeneity, ensuring communication efficiency, and scaling to support numerous devices.

2.2.2 Analysis of the Federated Learning ED Dataset

FL analyses datasets dispersed across several devices without centrally aggregating them, particularly in emotion recognition tasks. This method allows for cooperative model training while continuing data privacy. Assuring data consistency across heterogeneous sources, controlling device heterogeneity, and addressing potential biases generated by different data distributions are some of the key issues in analyzing such datasets. These difficulties are lessened by FL, which permits local model updates on every device. These changes are then combined to improve a global model without requiring the raw data to leave the devices. Preprocessing stages include data normalization, feature extraction and maybe data augmentation to improve model generalization in order to analyze such datasets successfully. Recurrent neural networks (RNNs) for sequential data, such as text or time-series physiological signals, or CNNs for image-based emotions are examples of ML models appropriate for FL in ED. To sum up, in order to analyze ED datasets for FL, it is necessary to undertake thorough preprocessing, choose suitable models, and pay close attention to privacy and performance measures that are specific to distributed learning settings.

2.2.3 Problems in FL Dataset Selection

Making sure datasets abide with privacy laws like the GDPR and getting consent from data owners to use and share their information safely in FL. In FL, striking a balance between a variety of representative, heterogeneous data sources and preserving data quality and schema uniformity throughout all involved clients. Collaborative learning: identifying and reducing biases to guarantee fair model performance across many demographic groups. Managing the communication overhead associated with FL as well as restrictions on the processing, memory, and storage capabilities of client devices. FL frequently involves handling identically dispersed and non-independent data in an efficient manner to provide reliable model training. Ensuring sure datasets may be split and distributed among numerous clients in an efficient manner while preserving successful training.

3 CONCLUSIONS

Choosing an appropriate dataset for ED is essential to developing dependable, widely applicable models that preserve anonymity. Datasets like ISEAR, IEMOCAP, RAVDESS, CREMA-D, EmoDB, CK+, FER-2013, MMI, JAFFE, AffectNet, SemEval-2018, and SEMAINE provide a great variety of modalities: text, audio, and facial expressions. These datasets form the basis for an in-depth study of emotion recognition. For example, more insightful audio samples into very emotional speech come forth from IEMOCAP and RAVDESS, whereas the ISEAR dataset is rich with textual data in terms of emotional responses. Again, some of these datasets are focused on facial expressions relevant to visual ED, such as FER-2013, CK+, and JAFFE. These datasets can be used by several clients due to the fact that FL is innately distributed, which will not risk user privacy and promote data diversity in order to increase model performance. Comprehensively including multi-modal datasets ensures an all-round approach in emotion identification to improve the accuracy of capturing the complexity of human emotions. However, class disparities should be dealt with and the databases should represent real situations. FL can significantly advance the science of ED by thoughtful selection and class balancing of these datasets, so that more morally correct applications be derived in health and other fields that are more personalized. Future applications of emotion detection with FL include promising enrichments of human-machine interactions and protections of the privacy of users. This develops in terms of multimodal integration, privacy, scalability, cultural diversity, and real-time applications.

REFERENCES

Adoma, A.F., Henry, N.M., Chen, W. and Andre, N.R., December. Recognizing emotions from texts using a bert-based approach. In 2020 IEEE 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTI P), (2020) 62-66.

AffectNet Dataset: https://www.kaggle.com/datasets/ngot hienphu/affectnet

Alotaibi, F.M., Classifying text- based emotions using logi stic regression. VAWKUM Transactions on Computer Sciences, 7(1), (2019) 31-37.

Asghar, M.Z., Subhan, F., Imran, M., Kundi, F.M., Shamshirband, S., Mosavi, A., Csiba, P. and Varkonyi-Koczy, A.R., Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content. arXiv preprint arXiv:1908.01587 (2019).

CK+ Dataset: https://www.kaggle.com/datasets/davilsena/ckdataset

CREMA- D Dataset: https://www.kaggle.com/datasets/ejl ok1/cremad

EmoDB Dataset: https://www.kaggle.com/datasets/piyush agni5/berlin-database-of-emotional-speech-emodb

FER2013 Dataset: https://www.kaggle.com/datasets/msa mbare/fer2013/code

G. K. J. Hussain and G. Manoj, Federated Learning: A Survey of a New Approach to Machine Learning, 2022 First International Conference on Electrical, Electronic s, Information and Communication Technologies (ICE EICT), Trichy, India, (2022) pp. 1-8.

IEMOCAP Dataset: https://www.kaggle.com/datasets/sa muelsamsudinng/iemocap-emotion-speech-database

JAFFE Dataset: https://www.kaggle.com/code/mpwolke/j apanese-female-facial-expression-tiff-images

Jain, S. and Asawa, K., Modeling of emotion elicitation conditions for a cognitive- emotive architecture. Cogn itive Systems Research, 55, (2019) 60-76.

MMI Dataset: https://kaggle.com/datasets/zaber666/meld-dataset

RAVDESS Dataset: https://www.kaggle.com/datasets/uw rfkaggler/ravdess-emotional-speech-audio