

UE20CS390A – End Semester Assessment

Speech Emotion Recognition through Federated Learning for Quality Assurance in call centers

Project ID : 85



Project Guide : Prof. Swathy M

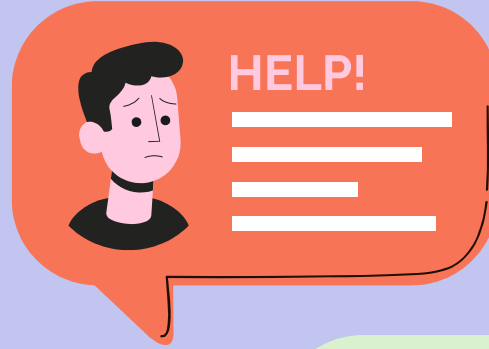
Project Team :

Andre Roy	- PES2UG20CS048
Ann Kurian	- PES2UG20CS055
Arshan Rodrigues	- PES2UG20CS067
Kristal D'souza	- PES2UG20CS170



Agenda

- Problem statement
- Scope
- Summary of Literature Survey
- Suggestions from Review – 2
- Proposed Methodology / Design Approach 
- Architecture
- Design Description
- Technologies Used
- Project Progress 
- References



Problem Statement



Current Agent evaluation

Bajaj Rural Penal					
S. No.	Agent Name	Target	Achieved	Achieved %	Yet To Do Count
1		7000	2466	35.2	4534
2		7000	912	13.0	6088
3		7000	1180	16.9	5820
4		7000	932	13.3	6068
5		7000	920	13.1	6080
6		7000	450	6.4	6550
7		7000	287	4.1	6713
8		7000	4793	68.5	2207
9		7000	450	6.4	6550
10		7000	4236	60.5	2764
11		7000	916	13.1	6084
12		7000	1223	17.5	5777
Grand Total		35000	18765	53.6	16235

Problem Statement



- Call centers receive a large number of customer calls on a daily basis, and it is important to ensure that these calls are of **high quality** and **customer satisfaction** is maintained.
- Assessing the quality of customer interactions with agents can be a challenging task as traditional methods of quality assurance rely on manual evaluation of calls, which can be **time-consuming** and **prone to errors**.
- Therefore, there is a need for an automated system that can accurately and efficiently evaluate customer interactions with agents.
- However, building such a system requires a large amount of labeled data, thereby bringing forward three main issues:
 - a. Difficult to obtain from a single call center
 - b. Difficulty in obtaining data from different call centers due to privacy concerns
 - c. Data storage

Proposed Solution



- **Speech Emotion Recognition (SER)** is technique that can be used to automatically analyze the emotional state of both the customer and the representative during the conversation.
- By analyzing the **acoustic and linguistic** features of speech, a system can identify emotions expressed by both customers and agents during the call.
- To address this problem, **Federated learning** can be used to train a speech emotion recognition model on data from **multiple call centers** without sharing the raw data or storing them on a single system.
- The model can then be used to analyse customer interactions with agents, providing call centers with valuable insights into the quality of their customer interactions as well as the performance of the customer service representatives.

Abstract

Emotion Recognition has a wide range of applications in call centers, including the avoidance of customer abuse, efficient matching of customers versus agents, etc.,

Federated Learning is a promising approach to address these challenges, as it enables the training of a highly accurate model while maintaining the privacy of agents and operational secrets. It can help call centers improve their customer experience and take their operations to the next level.



Literature Survey : Federated Learning

SNo.	Paper Details	Objectives of paper, Techniques/Methods	Advantages	Limitations
1.	Daniel Rueckert, Jonathan Passerat-Palmbach (2021) "Federated learning : Opportunities and challenges"	<ul style="list-style-type: none"> ▪ Aims to highlight the fields in which Federated learning can be applied and what possible applications they have. ▪ Also highlights the setbacks of this framework 	<ul style="list-style-type: none"> ▪ Saves space needed on central server as all the data is not stored on a single computer ▪ Provides means of collaborative ML while maintaining privacy and has better guarantee than other algorithms 	<ul style="list-style-type: none"> ▪ It is hard to identify malicious user as it is distributed ▪ Susceptible to attackers ▪ Communication bottleneck - can send relevant weights and not other information
2.	A Review of Applications in Federated Learning	<p>Deep Learning Models: The most commonly used deep learning models for federated learning Like CNN</p> <p>Differential Privacy: Differential privacy techniques are used to protect the privacy of user data</p> <p>Secure Multi-Party Computation (MPC)</p>	<p>Data Privacy: Federated learning ensures that user data remains on the device</p> <p>Improved Model Performance: Decentralized Training: Federated learning enables decentralized training, which reduces the reliance on a central server and makes the training process more efficient and scalable.</p>	<p>Heterogeneous Data: Federated learning works best when the data distribution among different devices is similar.</p> <p>Communication Overhead: Federated learning requires communication between the central server and the participating devices.</p> <p>Privacy Risks: The FL model can be susceptible to attackers</p>

Literature Survey : Speech-Emotion Recognition

SNo.	Paper Details	Objectives of paper, Techniques/Methods	Advantages	Limitations
3.	Płaza, Mirosław, et al. "Emotion Recognition Method for Call/Contact Centre Systems." Applied Sciences 12.21(2022): 10951.	The paper proposes an approach integrating data balancing, vectorization, 3 component modules (text and voice analyzer, Transcriptor) with classifiers such as SVM, ABC, DT, kNN, RFC, NBC for text and SVM, CNN, LDA and KNN for voice	<ul style="list-style-type: none"> ▪ Recognizing emotions in both text and voice channels with additional possibilities for using transcriptions of recordings ▪ Analysis of the effectiveness of classification depending on the types of classifiers used 	<ul style="list-style-type: none"> ▪ The model developed supports only Polish language ▪ The database used was not large enough leading to a less ideal results where average of metrics were 50-60% from text channels and 60-80% from voice channels
4.	Aashish Agarwal, Torsten Zesch (2019) "German End-to-end Speech Recognition based on DeepSpeech"	<ul style="list-style-type: none"> ▪ Making deepspeech better by changing hyper parameters by preprocessing. Using graphs like WER vs others ▪ Lesser WER by combining dataset when training and testing 	<ul style="list-style-type: none"> ▪ Does not explore effect of different recording environments or noise levels on emotion classification. 	<ul style="list-style-type: none"> ▪ Considers only accuracy as a performance measure ▪ Does not explore effect of different recording environments or noise levels on emotion classification.

Literature Survey : Speech-Emotion Recognition

SNo.	Paper Details	Objectives of paper, Techniques/Methods	Advantages	Limitations
5.	Abdelaziz A. Abdelhamid, "Robust Speech Emotion Recognition Using CNN+LSTM Based on Stochastic Fractal Search Optimization Algorithm", IEEE	<ul style="list-style-type: none"> Convolutional neural network(CNN) composed of four local feature-learning blocks and a long short-term memory (LSTM) layer for learning local and long-term correlations in the log Mel-spectrogram of the input speech samples 	<ul style="list-style-type: none"> Novelty: The SFS optimization algorithm used in the proposed approach is a novel approach that has not been widely used in speech emotion recognition research High accuracy: The proposed approach achieves high accuracy in recognizing emotions from speech signals 	<ul style="list-style-type: none"> The dataset used in the study is small, limiting the generalizability of the results. The system may not be able to recognize emotions in noisy environments or with poor quality audio.
6.	B. Li, D. Dimitriadis and A. Stolcke, "Acoustic and Lexical Sentiment Analysis for Customer Service Calls," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019	<ul style="list-style-type: none"> To analyze the contributions of acoustic and lexical features to the sentiment analysis system. Deep neural networks for acoustic feature extraction and classification Convolutional neural networks for lexical feature extraction and classification 	<ul style="list-style-type: none"> The system achieved a high accuracy in predicting the sentiment of customer service calls, which demonstrates its potential to be useful in real-world applications. The combination of acoustic and lexical features can provide a more comprehensive understanding of the sentiment 	<ul style="list-style-type: none"> The dataset used to evaluate the system may not be representative of all customer service calls The study does not consider the impact of speaker characteristics or contextual factors on the sentiment expressed in customer service calls.

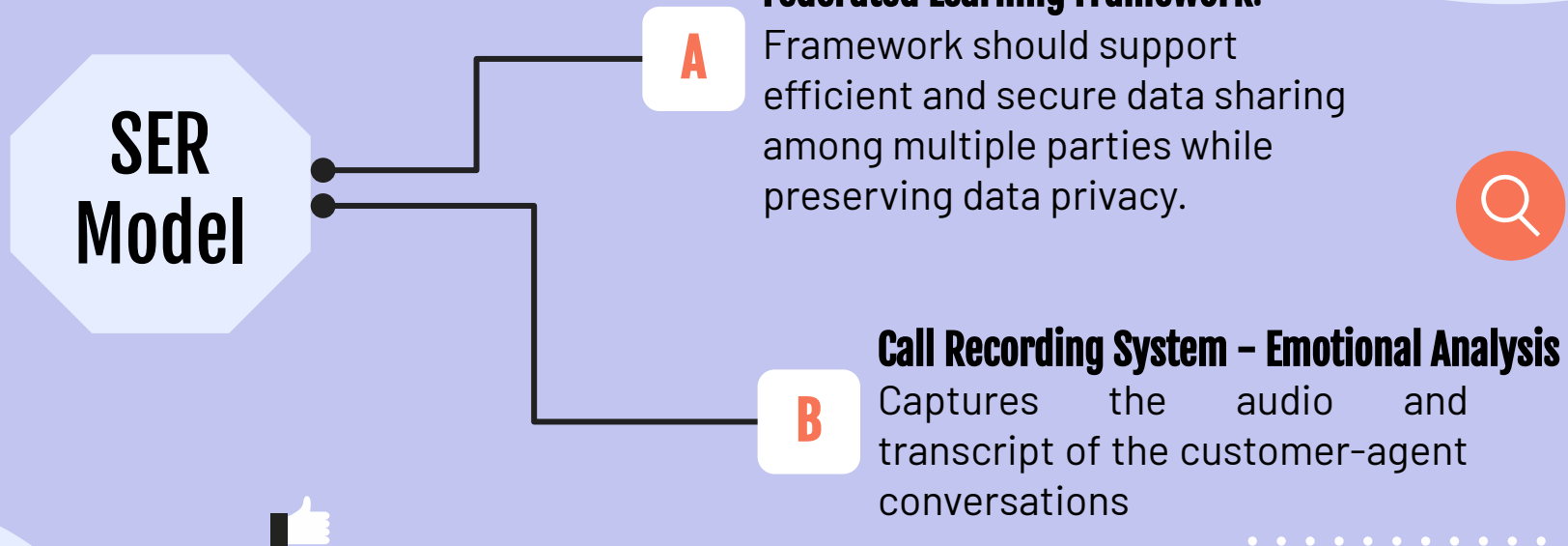
Summary of Literature Survey from Review 2

- **Small Sample Size** : Most of the research papers have used a very small dataset to evaluate their model. This limits the generalizability of the results
- **Limitations in Evaluation Metrics** : Majority of the research papers concentrates' on the basic evaluation metrics of ML and DL which fails to capture the full range performance of of their proposed work
- **Ethical Considerations** : The privacy and confidentiality of the call centre data are not adequately addressed in most of the research works
- **Emotional Analysis** : The existing research papers focuses on a single specific target emotion as output of the call recording analysis
- **Regional Challenge** : The existing research papers does not concentrate on Indian Accent based call recordings

Suggestions from Review 3

- **Suggestions on architecture diagram**
- **Dataset clarity**

Design Details



Design Details

Should be accurate and compatible with the chosen framework

Emotion Recognition API

1st



2nd



Tools such as TensorFlow and packages like Librosa for analyzing and processing audio data

Data processing

Federated learning to ensure adequate security and privacy of customer and agent data

Security & Privacy

3rd



Design Details

Depends on the quality of the speech data, complexity of the model, computing resources available, and the communication network's speed and reliability

Performance

4th



5th



System should be interoperable and communicate seamlessly with other systems using standard APIs and protocols

Interoperability

Capable of handling large volumes of audio data and can scale to meet the needs of growing call centers

Scalability

6th



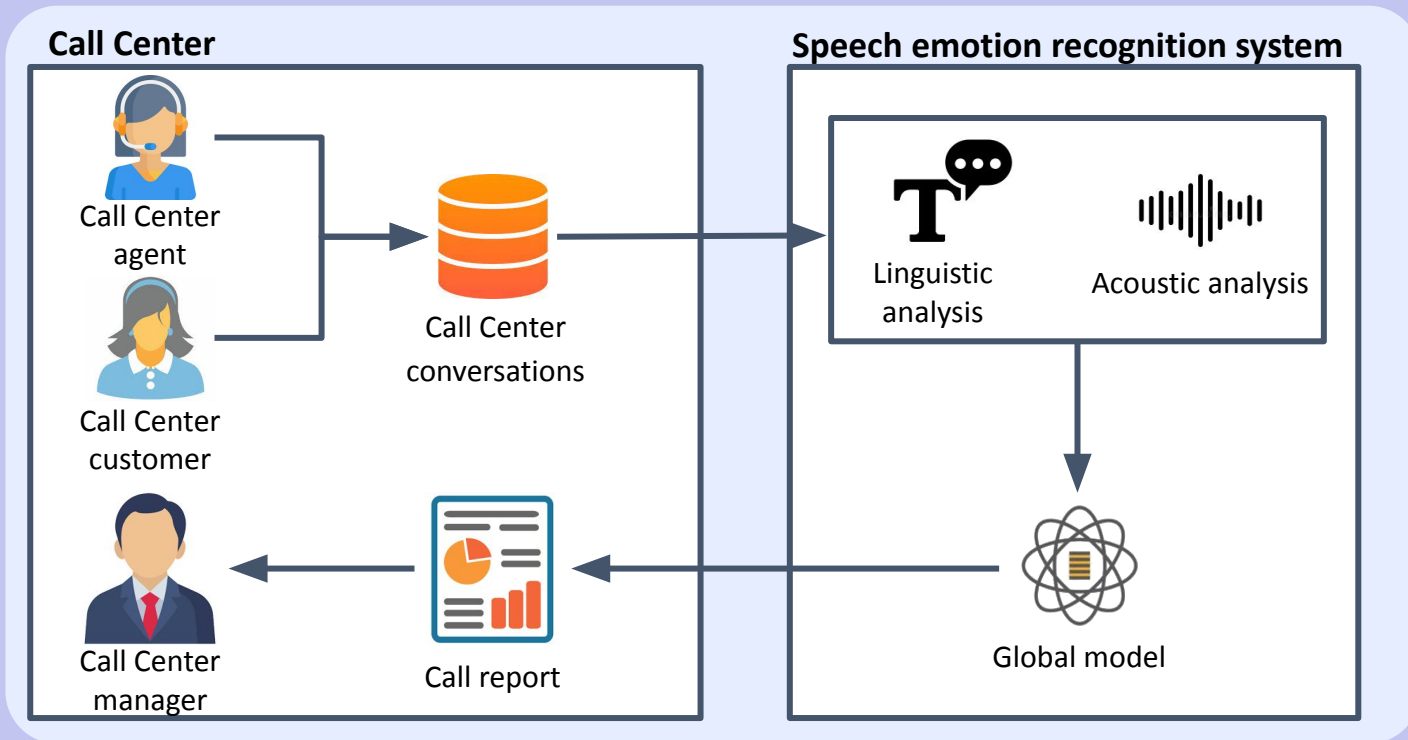
Design Constraints, Assumptions & Dependencies

- Data should be **labelled** for training the model
- The server must not be connected to the internet
- Customer and agent participation
- Results generated must be accessed only by manager/ admin and not every employee
- **Training data distribution** must be balanced
- Call center policies: The project assumes that call center policies are supportive of the use of emotion recognition technology, and that there are no legal or ethical constraints that would prevent its implementation.

Risks

- Security of highly sensitive data is important, hence our system has to be offline
- Final model might not be compatible with Indian English and names.
- Ethical consideration, potential impact on customers' privacy, dignity, and autonomy, and to ensure that the technology is used in a responsible and transparent manner
- The accuracy and effectiveness, impacted by factors such as the quality of the audio recordings, variations in accents and dialects, and the complexity of the emotions being analyzed

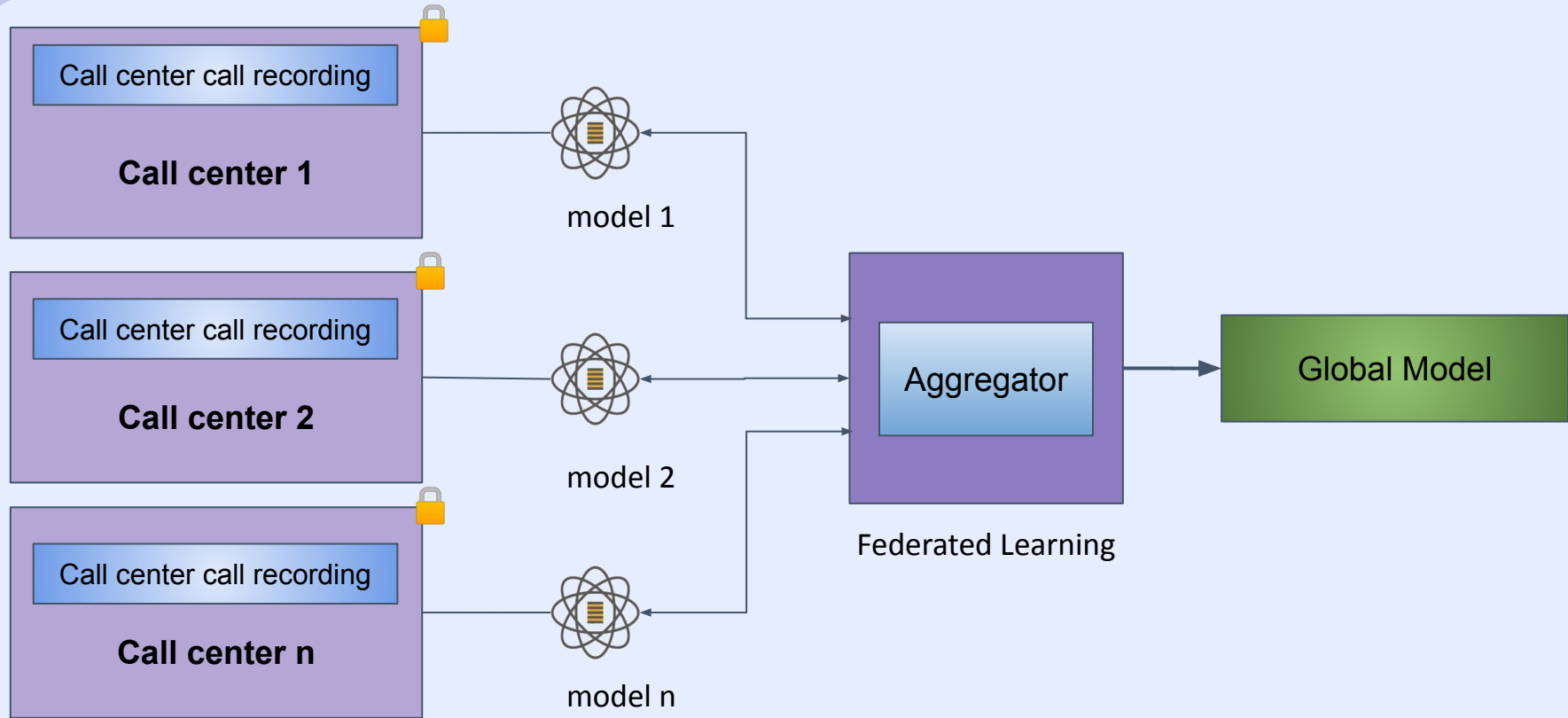
Architecture Diagram – Proposed Work



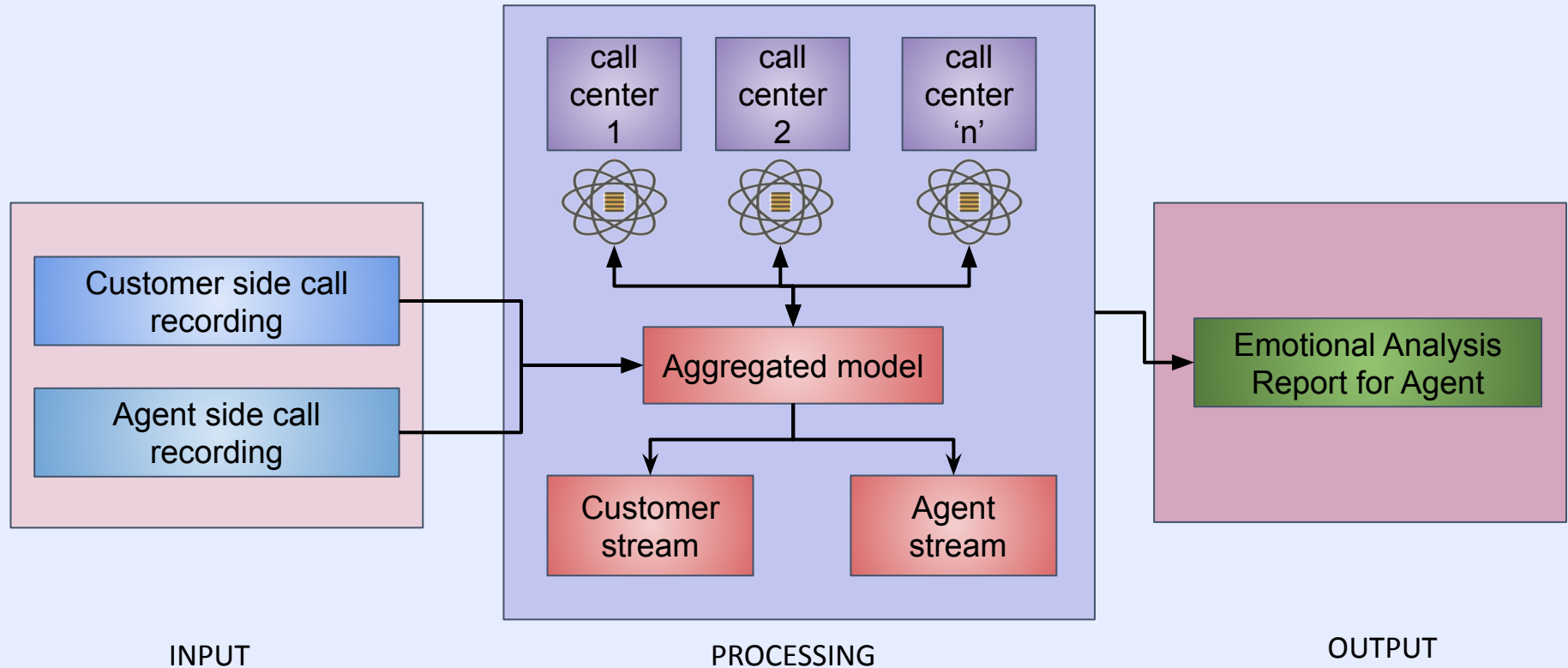
High Level Design – Proposed Work



Model training



High Level Design – Proposed Work



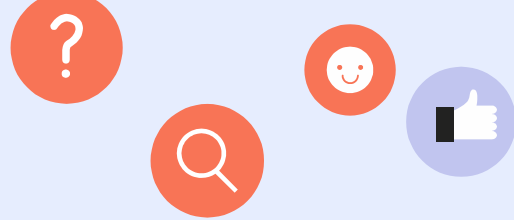
Proposed Methodologies

Data Annotation: Labelling of audio files with the emotion categories (happy, angry, sad, neutral) for both the customer and the agent

Data Processing: Data processing tools such as **TensorFlow** and packages like **Librosa** and **pyAudioAnalysis** for analyzing and processing audio data

Feature Extraction: Acoustic features including pitch and spectral characteristics of audio signal can be extracted using **Mel Frequency Cepstral Coefficient** or **Log-Mel spectrogram** while linguistic features including sentiment and syntax analysis can be done using **Wav2Vec model** or **Bert model**

Proposed Methodologies



Model Selection: A multi-modal model architecture that can combine both acoustic and linguistic features to predict emotion categories. We will experiment with different model architectures such as **Deep Emotion** (Convolutional Neural Network with LSTM, **Support Vector Machine** with optimization techniques) to find the one that performs best.

Deployment: Deploy the call center quality assurance system to a production environment. This may involve integrating with existing call center systems, and training call center managers on how to use the system.



Proposed Methodologies



Emotion Recognition API:

- Should be accurate and compatible with the chosen framework

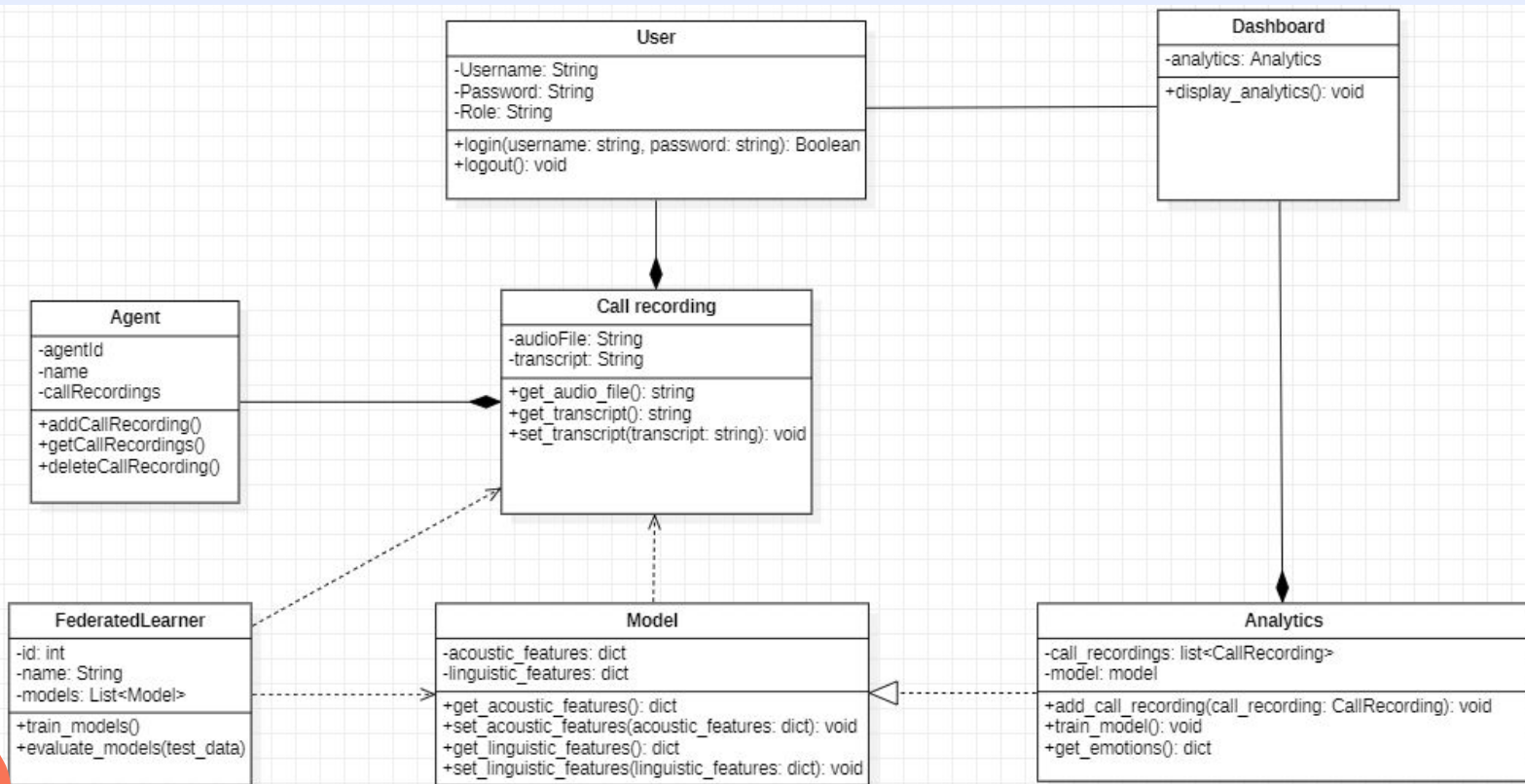
Data Processing:

- Data processing tools such as **TensorFlow** and packages like **Librosa**, **pyAudioAnalysis** for analyzing and processing audio data

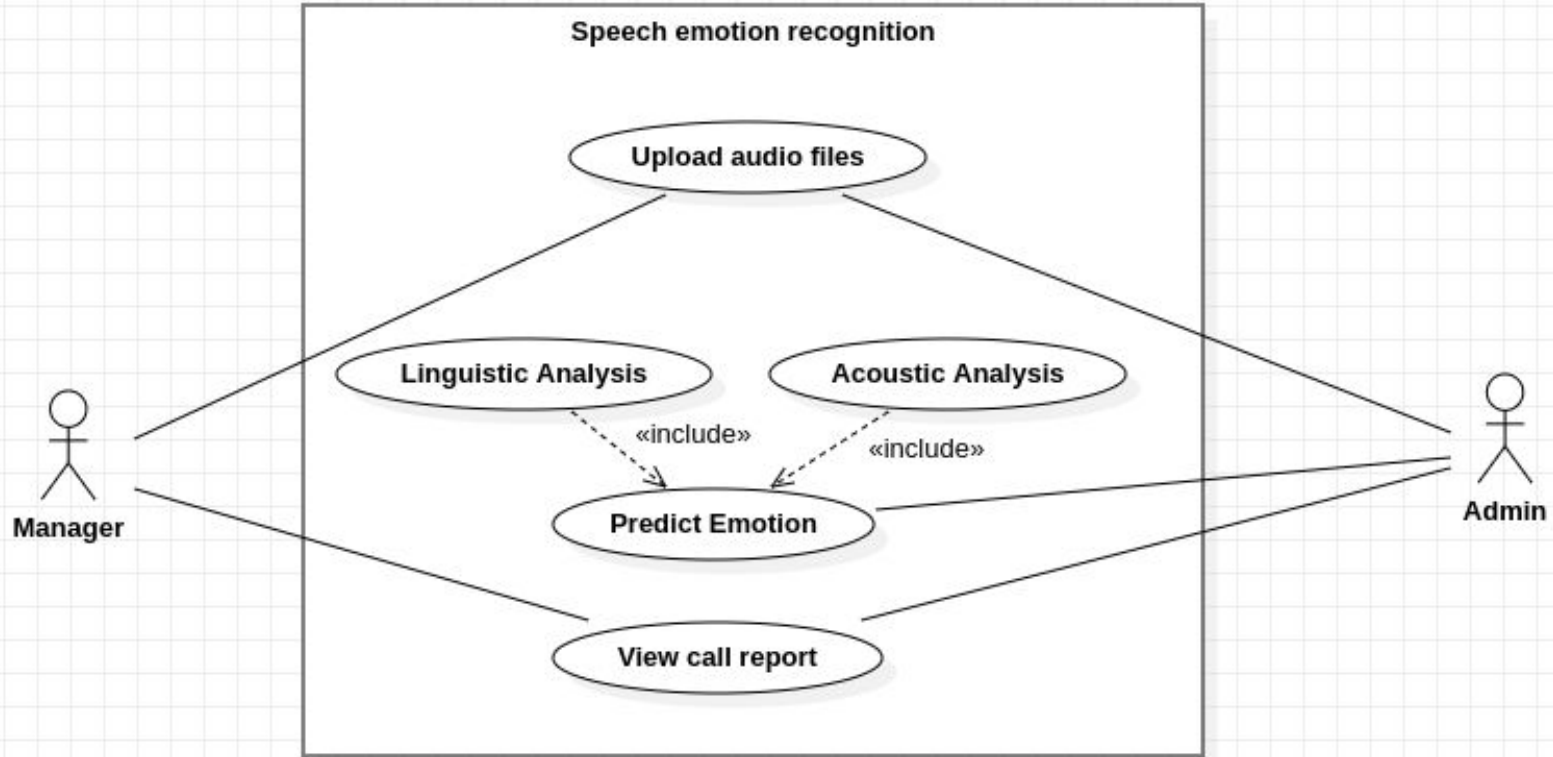
Models Selection:

- **CNN + LSTM** with **Log-Mel Spectrogram**, **MFCC**, **MFMC**
- **SVM classifier** for **MFMC**

Master Class Diagram



Use Case Diagram



User interface

Agent Analysis Dashboard

Home Call Analytics Agent Performance Emotion Detection

Agent Performance Analysis

Date	Duration	Satisfaction Score	Emotion Detected	Metric#3	Metric#4	Metric#5
2023-04-30	6 min	4.5	Happy	150 WPM	85%	70 dB
2023-04-29	4 min	3.5	Neutral	125 WPM	80%	65 dB
2023-04-27	5 min	4.0	Sad	135 WPM	90%	75 dB

Emotion Detection Analysis



Emotion Range Analysis



Project Progress



**Feasibility analysis
Of previous title**

January



**Data collection and
Literature Survey**

February

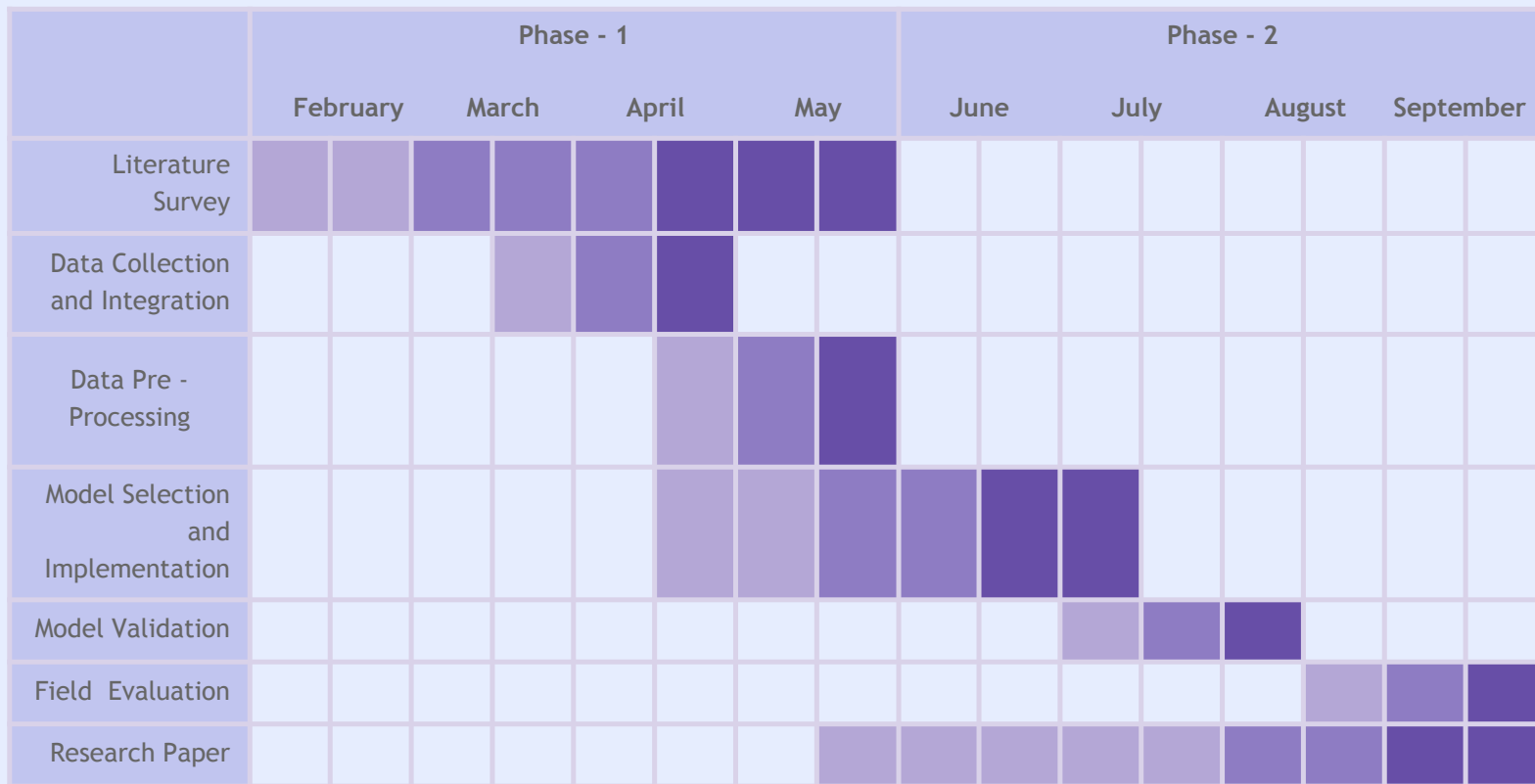


**Denoising and Pre
processing of audio files**

March



Gantt Chart





Conclusion



- We aim to provide an alternative to the manager who previously had to listen in on calls to evaluate performance of the agents.
- Instead our application would provide a simple interface which would show the analysis of the calls performed by the agent
- We proposed a privacy-preserving (Speech Emotion Recognition) SER model by utilizing Federated Learning to eliminate the assumption of abundant data availability on a single device.



References

- [1] Płaza, Mirosław, et al. "*Emotion Recognition Method for Call/Contact Centre Systems.*" Applied Sciences 12.21(2022): 10951. *
- [2] Potdar, Veena & Santhosh, Lavanya & Bhatt, Supritha. (2021). "*Analysis of Vocal Pattern to Determine Emotions using Machine Learning.*"
- [3] Petrushin, Valery. (2000). "*Emotion in Speech: Recognition and Application to Call Centers.*" Proceedings of Artificial Neural Networks in Engineering.
- [4] B. Li, D. Dimitriadis and A. Stolcke, "*Acoustic and Lexical Sentiment Analysis for Customer Service Calls,*" ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019
- [5] Daniel Rueckert, Jonathan Passerat-Palmbach "*Federated learning : Opportunities and challenges*" arXiv:2101.05428v1 [cs.LG] 14 Jan 2021.
- [6] Aashish Agarwal, Torsten Zesch (2019) "*German End-to-end Speech Recognition based on DeepSpeech*"
- [7] Yangyang Xia, Li-Wei Chen, Alexander Rudnicky, Richard M. Stern, INTERSPEECH 2021: "*Temporal Context in Speech Emotion Recognition*"

References

- [8] J. Ancilin, A. Milton, *"Improved speech emotion recognition with Mel frequency magnitude coefficient"*
- [9] Souraya Ezzat, Neamat El Gayar, and Moustafa M. Ghanem, *"Sentiment Analysis of Call Centre Audio Conversations using Text Classification"*, International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 4 (2012) pp. 619 – 627
- [10] Rashid Jahangir, Ying Wah Teh, Faiqa Hanif & Ghulam Mujtaba, *"Deep learning approaches for speech emotion recognition: state of the art and research challenges"*, Springer Science+Business Media, LLC, part of Springer Nature 2021, corrected publication 2021
- [11] I. Shafran and M. Mohri, *"A comparison of classifiers for detecting emotion from speech,"* Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., Philadelphia, PA, USA, 2005, pp. I/341-I/344 Vol. 1, doi: 10.1109/ICASSP.2005.1415120.
- [12] S. Yoon, S. Byun and K. Jung, *"Multimodal Speech Emotion Recognition Using Audio and Text,"* 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 2018, pp. 112-118, doi: 10.1109/SLT.2018.8639583.
- [13] Valery Petrushin, *"Emotion in Speech: Recognition and Application to Call Centers"*, Article · January 2000,

References

- [14] Blumentals, Eduards, and Askars Salimbajevs. *"Emotion recognition in real-world support call center data for latvian language."* CEUR Workshop Proceedings. Vol. 3124. 2022.
- [15] Li Lia,b, Yuxi Fana, Mike Tsec, Kuo-Yi Lina,b, *"A review of Applications in Federated Learning"* Elsevier – Computers & Industrial Engineering Volume 149, November 2020, 106854.
- [16] Abbaschian BJ, Sierra-Sosa D, Elmaghraby A. *"Deep Learning Techniques for Speech Emotion Recognition, from Databases to Models. Sensors."* 2021; 21(4):1249.
- [17] Byun S-W, Kim J-H, Lee S-P. *"Multi-Modal Emotion Recognition Using Speech Features and Text-Embedding."* Applied Sciences. 2021; 11(17):7967.
- [18] S. Lugović, I. Dunder and M. Horvat, *"Techniques and Applications of Emotion Recognition in Speech"*, MIPRO 2016, May 30 – June 3, 2016, Opatija, Croatia
- [19] Anna Bogdanova; Nii Attoh-Okine, F.ASCE; and Tetsuya Sakurai, *"End-to-end speech emotion recognition: challenges of real-life emergency call centers data recordings"* ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering Vol. 6, Issue 3 (September 2020)

? **THANK YOU**

