Acoustic and Linguistic Analysis for Early Dementia Detection using Machine Learning

Marcel Seitz & Philip Werz

02.06.2025, Kick-off Guided Research

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Who we are

Marcel Seitz

Philip Werz

Agenda

1. Motivation and Relevance
2. Research Gap & Research Questions
3. Theoretical Background
4. Methodology
5. Outlook
6. Timeline

1. Motivation and Relevance

Dementia

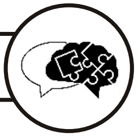
Challenges of Dementia

- Rising global prevalence of dementia, especially Alzheimer's Disease
- AD affected ~57 million in 2021, ~10 million new cases/year
- Major global cause of death & disability
- Urgent need for early diagnosis & long-term monitoring

ALPHA-KI

- AI-Based Digital Health Assistant for Elderly & Chronically Ill

AssistD



- Improving Dementia Care through Adaptive Voice Assistance

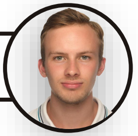
Early Markers and Assessment Tools

- Subtle changes in speech and language are often earliest detectable markers of cognitive decline

Cookie Theft Picture

- Widely used neuropsychological instrument

Free Speech



- Naturalistic voice assistant interactions

Acoustic and Linguistic Analysis using Machine Learning

Sources: [1-4]

2. Research Gap & Research Questions

Research Gap

ADReSS Challenge

- **AD classification:** Building a model to predict whether a speech session indicates AD or non-AD
- **MMSE score regression:** Creating a model to infer a subject's Mini Mental State Examination (MMSE) score from speech/language data

Few studies focus on HC vs. MCI; mostly focused on HC vs. AD/DM

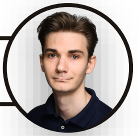
Lack of Multimodal Analysis

- Previous studies mostly use either acoustic (audio) or linguistic (text) features, sometimes both, but treat them separately.
- [5] focused on multilingual analysis on unified conversational dataset

Combining acoustic and linguistic features via <u>multimodal</u> ML remains largely unexplored

Research Questions

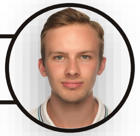
Cookie Theft Picture



I How accurately can ML models based on acoustic and linguistic features from **Cookie Theft Picture descriptions** distinguish between HC and MCI patients?

II Does combining audio and text features from Cookie Theft descriptions enhance early dementia classification compared to single-modality models?

Free Speech



I How well can ML models using acoustic and linguistic features from **Alexa conversations** differentiate HC and DM/MCI patients?

II Does integrating audio and linguistic features from Alexa-based free speech responses improve early dementia classification compared to unimodal approaches?

Sources: [5-6]; Abbreviations: AD = Alzheimer’s disease, MCI = Mild Cognitive Impairment, DM = Dementia, HC = Healthy Control

3.1 Theoretical Background

DM vs MCI

Clinical Background

- **Dementia**: Progressive cognitive decline affecting daily life; AD is the most common cause
- **MCI**: Intermediate stage between normal aging and DM; may progress, stay stable, or revert.
- **Early Detection**: Critical at MCI stage for timely intervention and treatment trials.

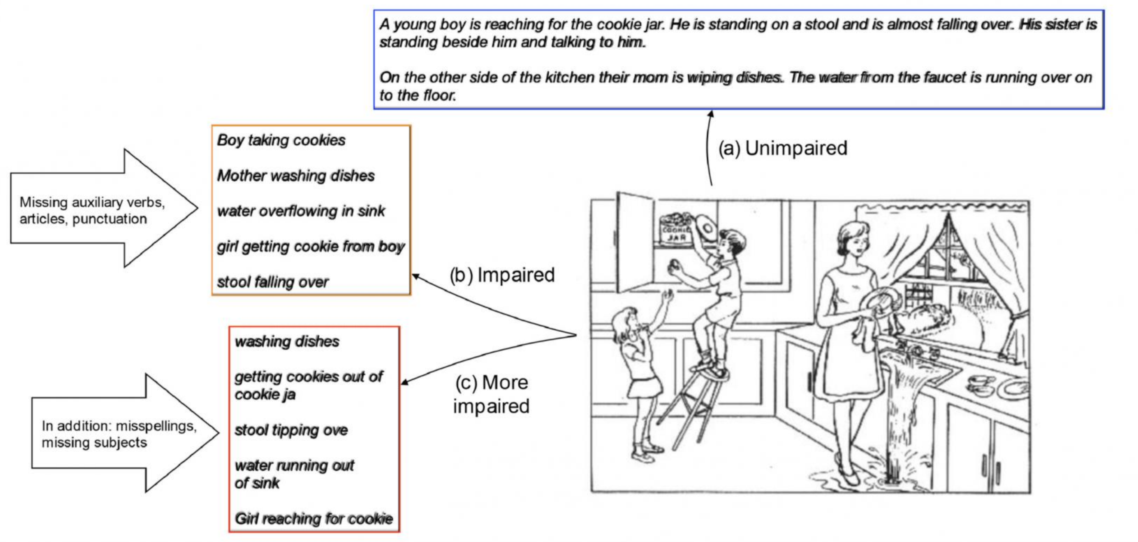
Early Biomarkers of Cognitive Decline

- **Speech as Cognitive Marker**: Involves memory, attention, and language, sensitive to early decline.
- **Early Language Changes in MCI**: Reduced lexical variety, vague words, and semantic errors.
  - **Simpler Syntax**: Shorter sentences, fewer embedded clauses.
  - **Fluency & Prosody**: More pauses, hesitations, and reduced articulation/emotional tone.

Speech

Standardized (Cookie Theft Picture)

- **Standardized tasks** (e.g., Cookie Theft) guide content and reduce variability, enabling easier feature extraction
- **Controlled & Natural**: Same visual stimulus enables comparability
- **Cognitive Support**: Reduces memory/attention demands, suitable for impaired participants
- **MCI Indicators**: Less detail, low lexical/syntactic complexity, more pronouns, reduced fluency



Sources: [7-21], [27]; Abbreviations: AD = Alzheimer's disease, MCI = Mild Cognitive Impairment, DM = Dementia, HC = Healthy

Source: Figure 1

3.2 Theoretical Background

DM vs MCI

Clinical Background

- **Dementia**: Progressive cognitive decline affecting daily life; AD is the most common cause
- **MCI**: Intermediate stage between normal aging and AD; may progress, stay stable, or revert.
- **Early Detection**: Critical at MCI stage for timely intervention and treatment trials.

Early Biomarkers of Cognitive Decline

- **Speech as Cognitive Marker**: Involves memory, attention, and language, sensitive to early decline.
- **Early Language Changes in MCI**: Reduced lexical variety, vague words, and semantic errors.
  - **Simpler Syntax**: Shorter sentences, fewer embedded clauses.
  - **Fluency & Prosody**: More pauses, hesitations, and reduced articulation/emotional tone.

Speech

Source: Figure 2

Free Speech (Amazon Alexa)

- **Free speech** (e.g., Alexa) offers more natural data but introduces higher linguistic and acoustic variability, posing challenges for consistent classification
- **Linguistic Markers**: Less coherence, fewer content words, more pronouns & pauses
- **Temporal Features**: Slower tempo, more hesitations, altered articulation

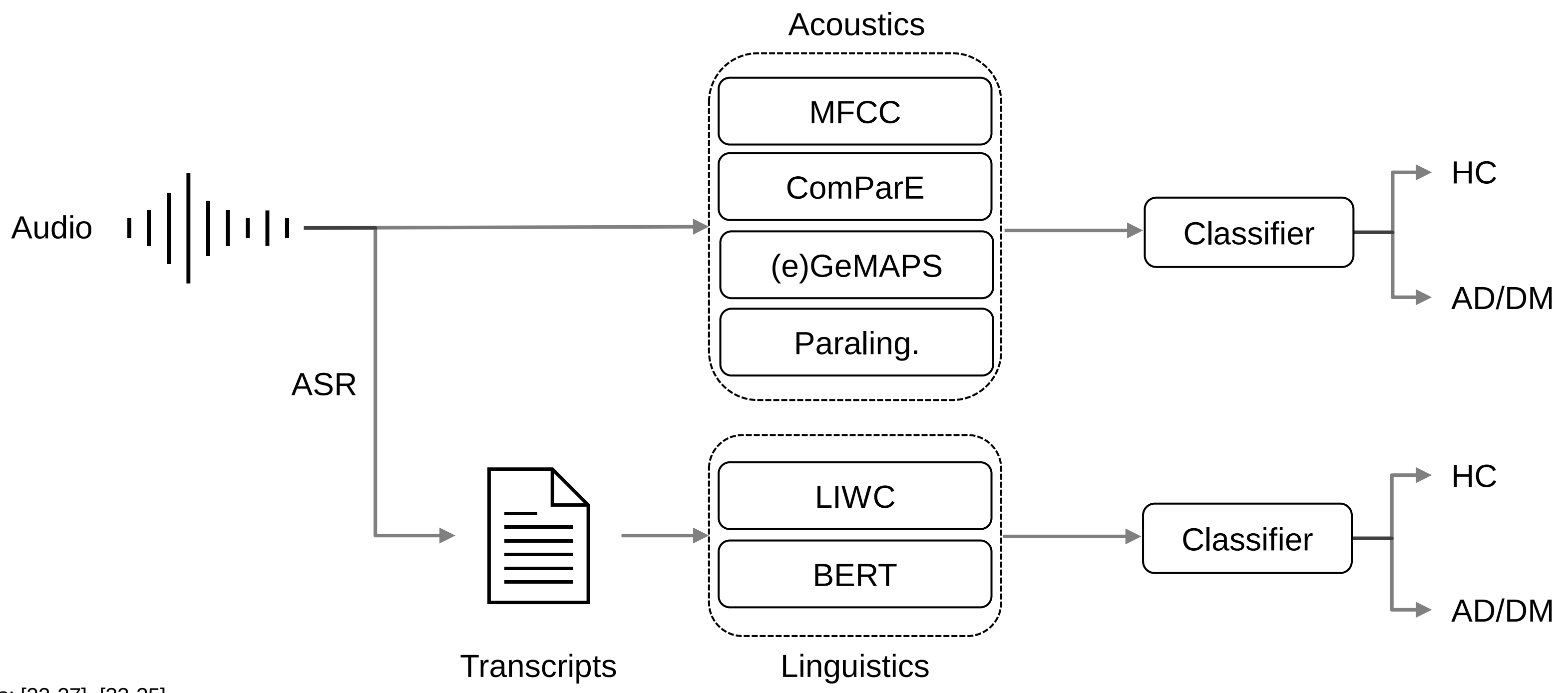
Sources: [7-10], [17-21], [27-32]; Abbreviations: AD = Alzheimer's disease, MCI = Mild Cognitive Impairment, DM = Dementia, HC = Healthy Control

3.3 Theoretical Background

Machine Learning

Cognitive Impairment Detection

- **ML detects cognitive decline**: Scalable, non-invasive alternative to MRI and lab tests
- **Deep learning**: CNNs, RNNs, and Transformers (e.g., BERT, wav2vec2) learn directly from speech/text
- **Key features**: MFCCs, eGeMAPS etc. reflect pitch and pauses



Legend

- **ASR**: Automatic Speech Recognition
- **Acoustics** : Acoustic feature extraction
- **Linguistics**: Text feature extraction

Sources: [22-27], [33-35]

3.3 Theoretical Background

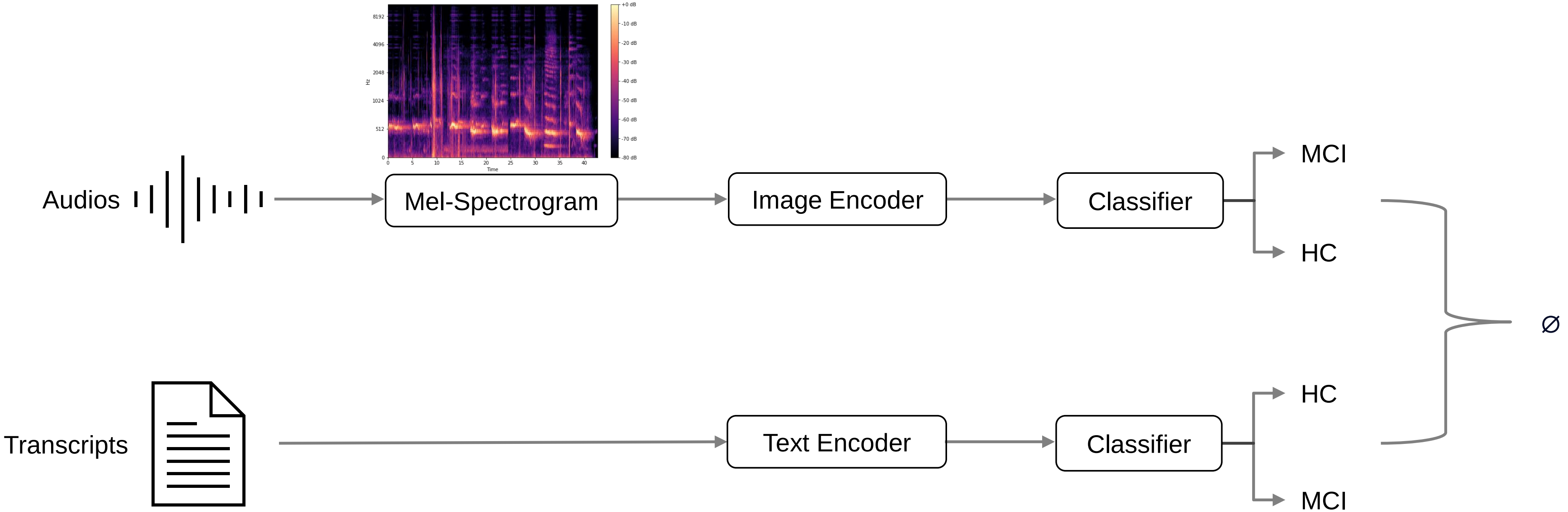
Machine Learning

Cognitive Impairment Detection

- **ML detects cognitive decline**: Scalable, non-invasive alternative to MRI and lab tests
- **Deep learning**: CNNs, RNNs, and Transformers (e.g., BERT, wav2vec2) learn directly from speech/text
- **Key features**: MFCCs, eGeMAPS etc. reflect pitch and pauses

Multimodal Fusion

- **Multimodal Fusion**: Combines speech and text for better dementia detection
  - **Late-Level Fusion:** Combines modalities after individual classification; processes each modality separately



Sources: [22-27], [33-35], Mel-Spectrogram: Figure 3

3.3 Theoretical Background

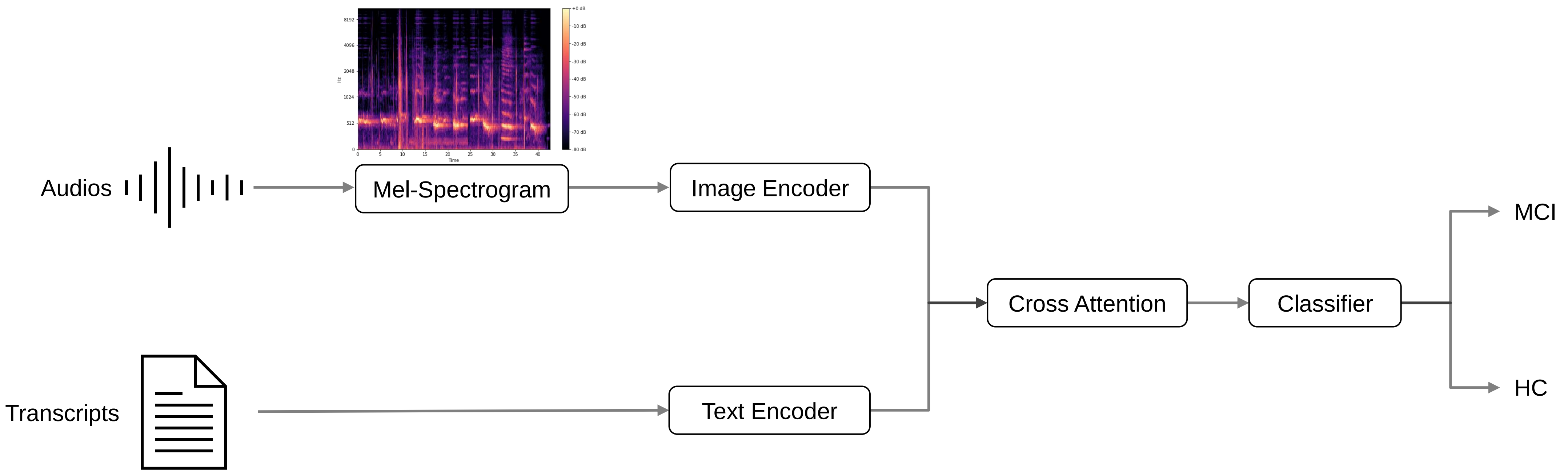
Machine Learning

Cognitive Impairment Detection

- **ML detects cognitive decline**: Scalable, non-invasive alternative to MRI and lab tests.
- **Deep learning**: CNNs, RNNs, and Transformers (e.g., BERT, wav2vec2) learn directly from speech/text.
- **Key features**: MFCCs, eGeMAPS etc. reflect pitch and pauses

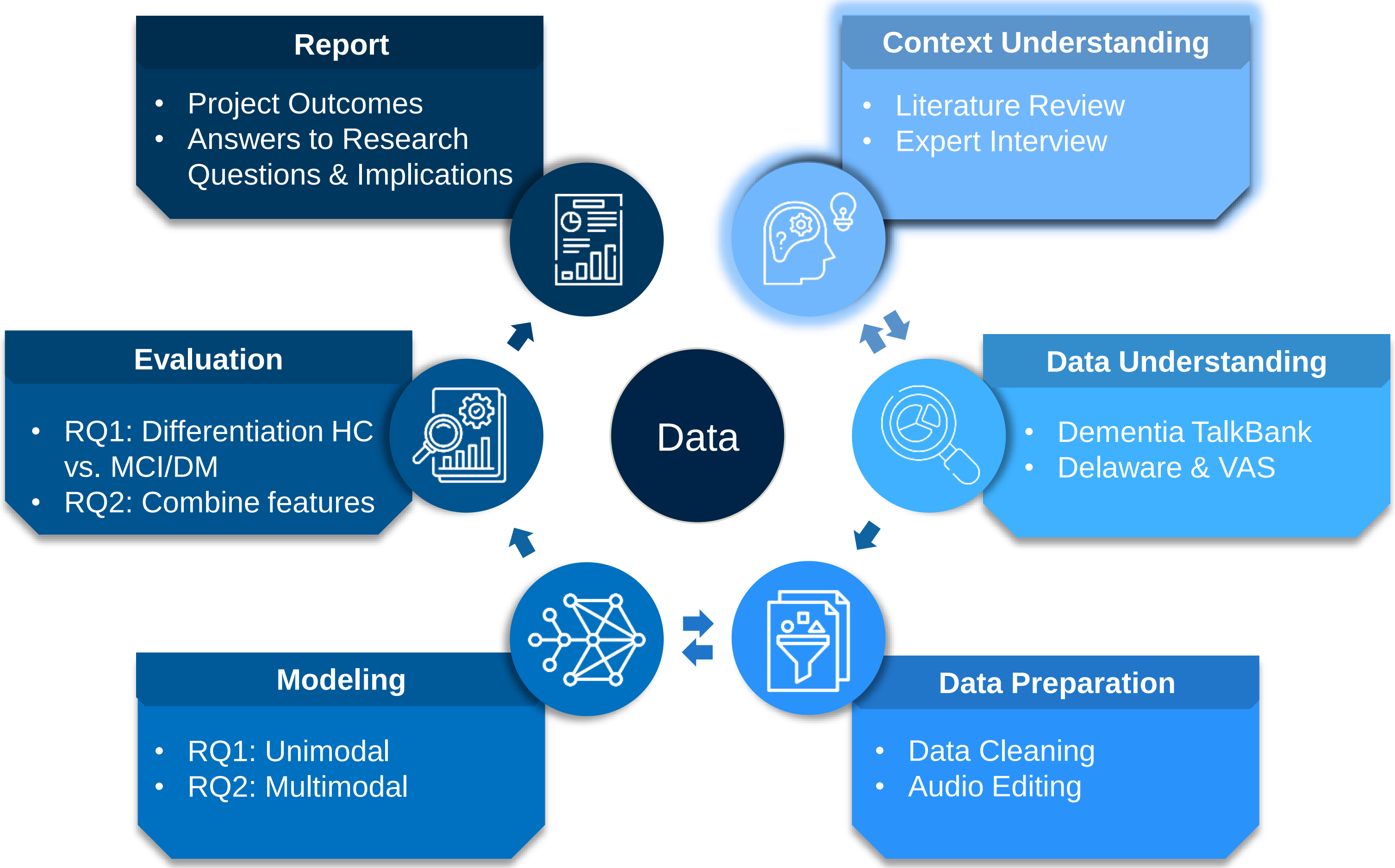
Multimodal Fusion

- **Multimodal Fusion**: Combines speech and text for better dementia detection.
  - **Early-Level:** Combines modalities before classification; merges features or embeddings; enables cross-modal interactions; requires synchronized inputs
    - **Cross-Attention:** One modality attends to another; captures inter-modal dependencies

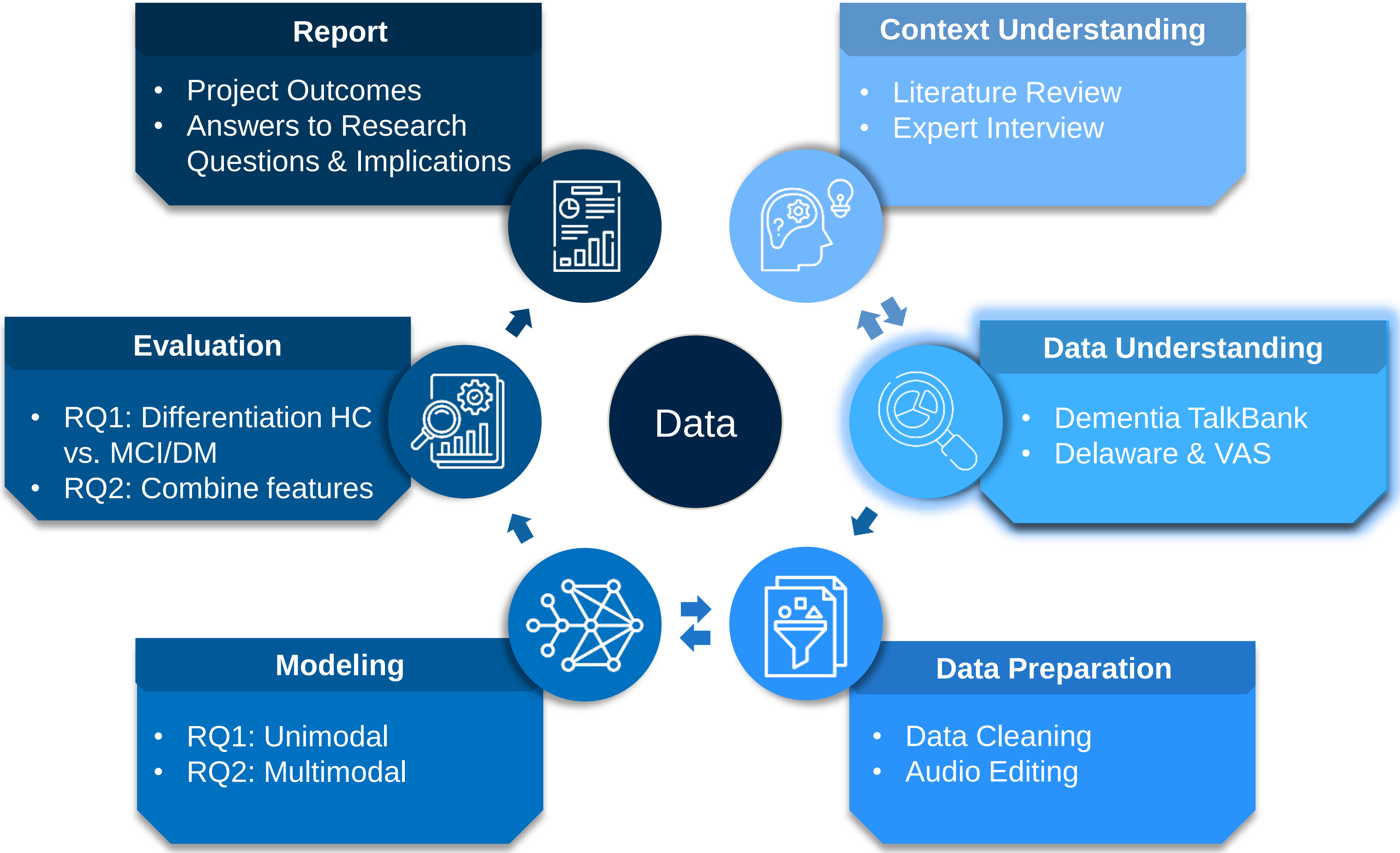


Sources: [22-27], [33-35], Mel-Spectrogram: Figure 3

4.1 Methodology – Context Understanding



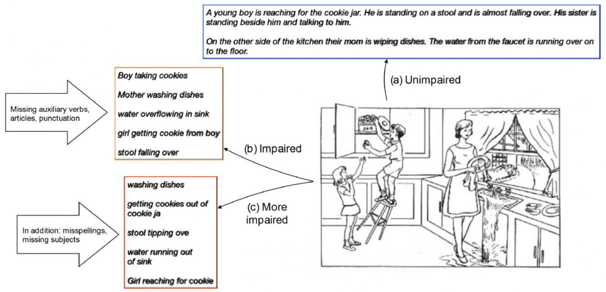
4.2 Methodology – Data Understanding



Source: [6], Abbreviations: MCI = Mild Cognitive Impairment, HC = Healthy Control

Delaware

- Standardized dataset
- Picture descriptions; Story-, Procedural-, Personal-Narratives
- Participants: older adults (61-91)
- Audio recordings and manually transcribed and annotated speech data (CHAT format, 95 Participants).
- Includes data from a control group and individuals diagnosed with MCI (61 MCI, 34 HC).
- Number of .mp3 files: 111
- Total Length: 18 h
- Data Collection Ongoing



Source: Figure 1

4.2 Methodology – Data Understanding

Report
- Project Outcomes
- Answers to Research Questions & Implications

Context Understanding
- Literature Review
- Expert Interview

Evaluation
- RQ1: Differentiation HC vs. MCI/DM
- RQ2: Combine features

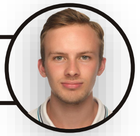
Data

Data Understanding
- Dementia TalkBank
- Delaware & VAS

Modeling
- RQ1: Unimodal
- RQ2: Multimodal

Data Preparation
- Data Cleaning
- Audio Editing

VAS

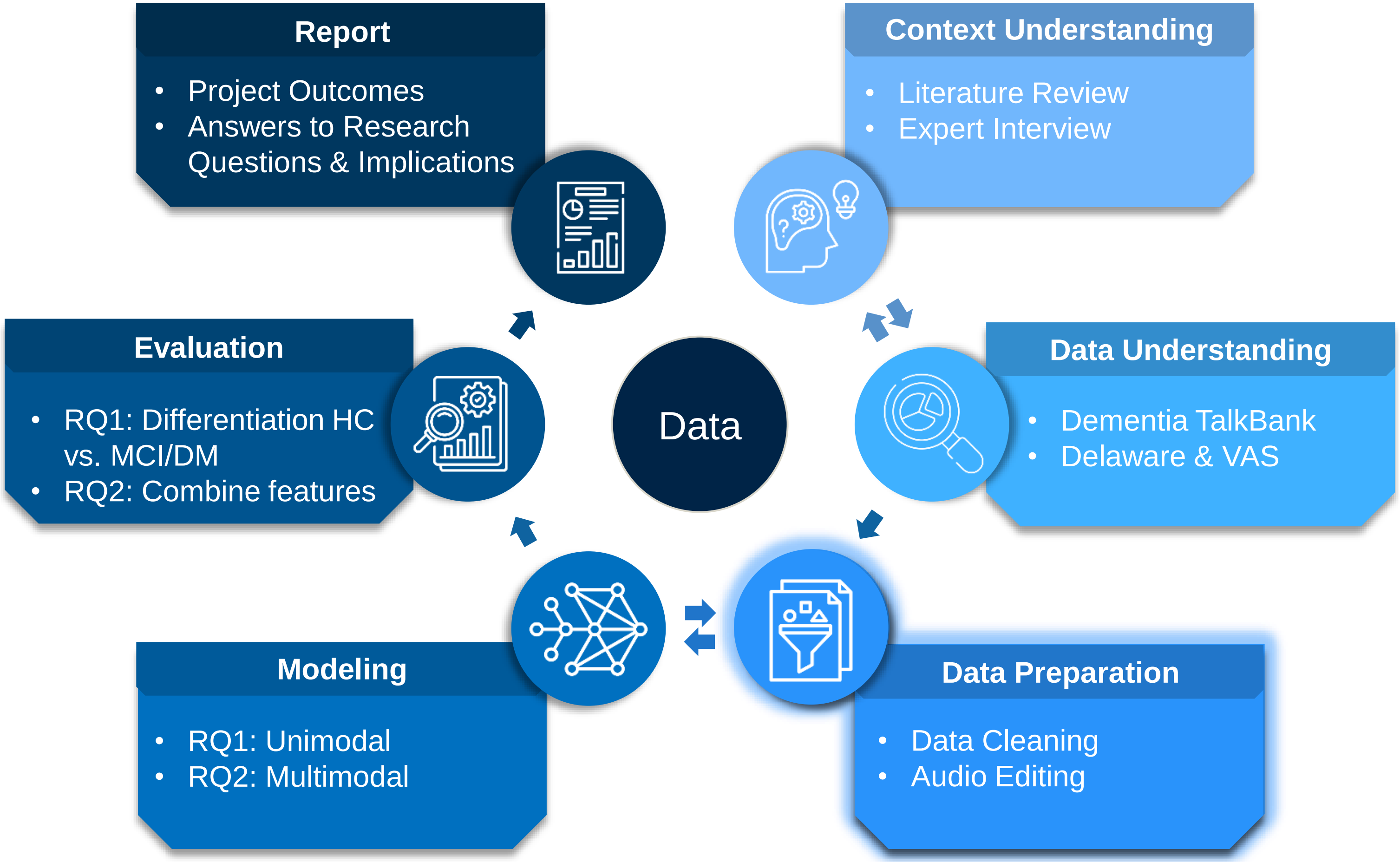


- Dataset of voice interactions from older adults (64-94).
- Audio recordings and text transcripts from Amazon Alexa (CHAT format, 101 Participants).
- Includes data from a control group and individuals diagnosed with dementia (DM:30, HC: 36, MCI: 35).
- Number of .mp3 files: 101
- Total Length: 3.5h

Source: Figure 2

Source: [6]; Abbreviations: MCI = Mild Cognitive Impairment, DM = Dementia, HC = Healthy Control

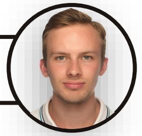
4.3 Methodology – Data Preparation



Delaware

- Clean Dataset
- Isolate participant speech
- Reduce to CTP descriptions
- Estimated Total Length: 3h

VAS



- Clean Dataset
- Split Audios
  - Separate audios for each Alexa request

Abbreviations: CTP = Cookie Theft Picture

4.4 Methodology – Modeling

Report
- Project Outcomes
- Answers to Research Questions & Implications

Context Understanding
- Literature Review
- Expert Interview

Evaluation
- RQ1: Differentiation HC vs. MCI/DM
- RQ2: Combine features

Data

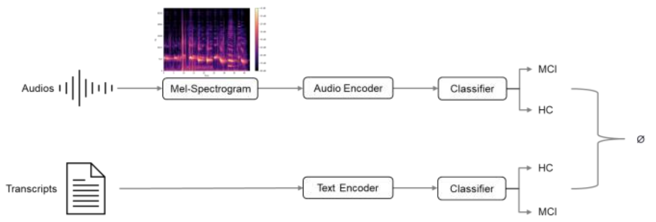
Data Understanding
- Dementia TalkBank
- Delaware & VAS

Modeling
- RQ1: Unimodal
- RQ2: Multimodal

Data Preparation
- Data Cleaning
- Audio Editing

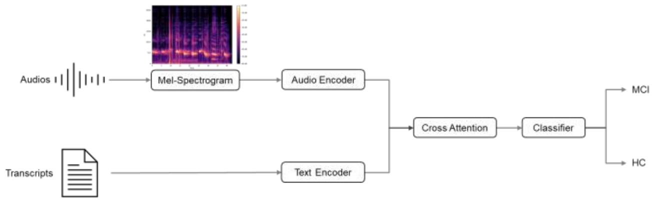
RQ1 - Unimodal

- Late Fusion
- Simplified Model:



RQ2 - Multimodal

- Early Fusion with Cross Attention
- Simplified Model:



Source: [Figure 3]

4.5 Methodology – Evaluation

Report
- Project Outcomes
- Answers to Research Questions & Implications

Context Understanding
- Literature Review
- Expert Interview

Data Understanding
- Dementia TalkBank
- Delaware & VAS

Data Preparation
- Data Cleaning
- Audio Editing

Modeling
- RQ1: Unimodal
- RQ2: Multimodal

Evaluation
- RQ1: Differentiation HC vs. MCI/DM
- RQ2: Combine features

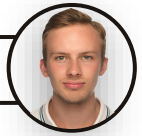
Data

Abbreviations: MCI = Mild Cognitive Impairment, DM = Dementia, HC = Healthy Control

Delaware

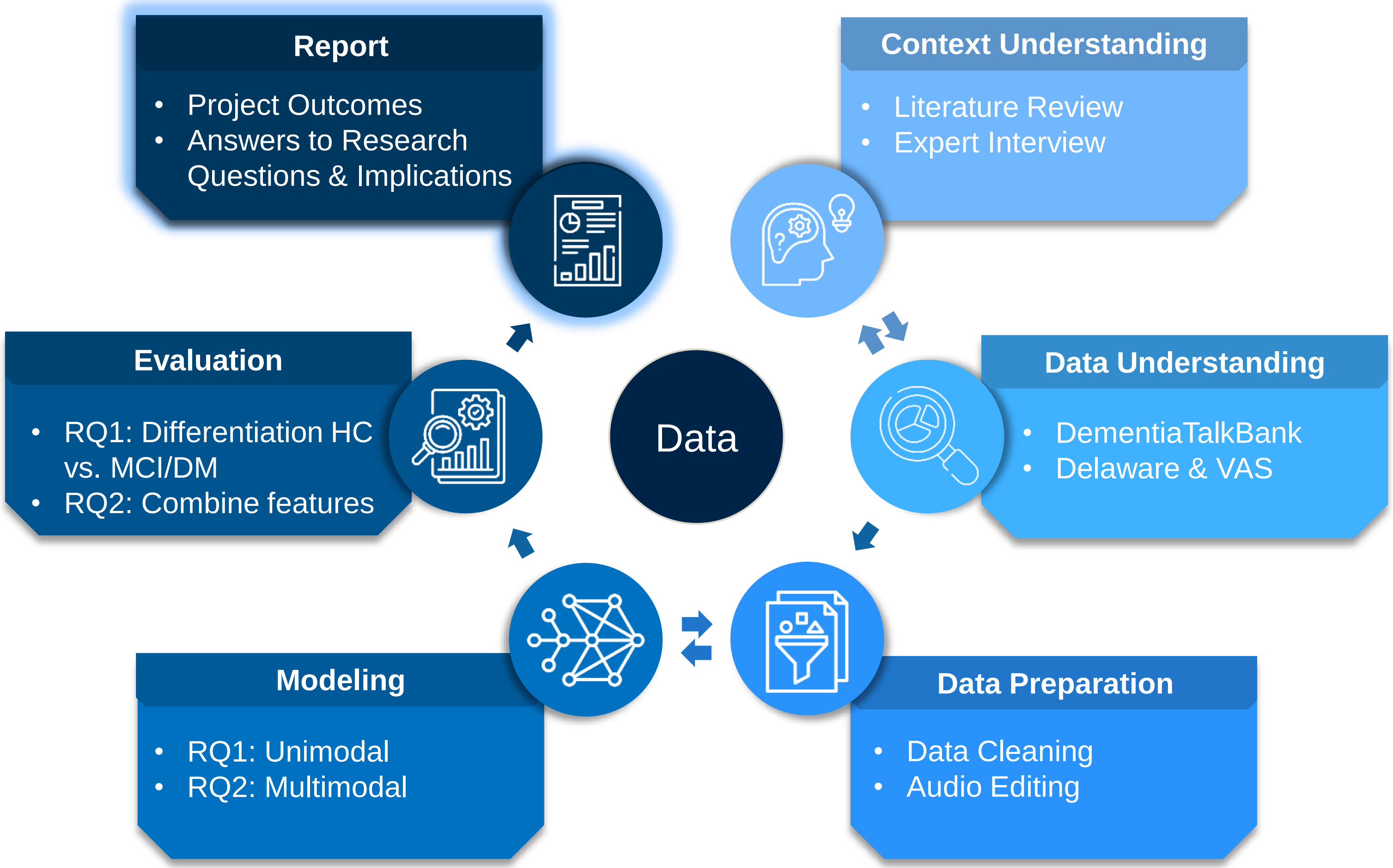
- How accurately can ML models based on acoustic and linguistic features from Cookie Theft Picture descriptions **distinguish** between HC and MCI patients?
- Does combining audio and text features from Cookie Theft descriptions enhance early dementia classification compared to single-modality models?

VAS

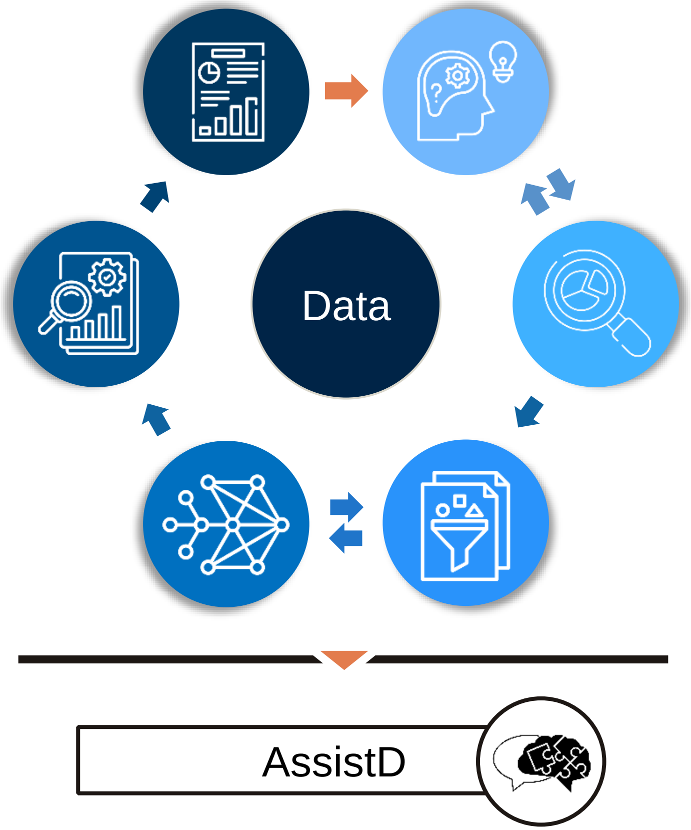


- How well can ML models using acoustic and linguistic features from Alexa conversations **differentiate** HC and DM/MCI patients?
- Does integrating audio and linguistic features from Alexa-based free speech responses improve early dementia classification compared to unimodal approaches?

4.6 Methodology – Report



4.7 Methodology – Future Work



Source: [4]

5. Outlook

Expected Contribution

Academic Insights

- **Deeper insights** into MCI vs. HC, addressing the prior focus on AD/DM vs. HC in most studies
- **Comparison** of multimodal vs. unimodal model performance
- **Comparative analysis** of spontaneous speech versus standardized speech in terms of classification accuracy

Practical Guidance

- **Test both standardized** (CTP) and **free/spontaneous** (Alexa) improves clearness in regard to future projects (e.g. Smartwatch)
- **Offer guidance** on (standardized) data collection and task setup for consistent dataset quality and better accuracy

Next Steps

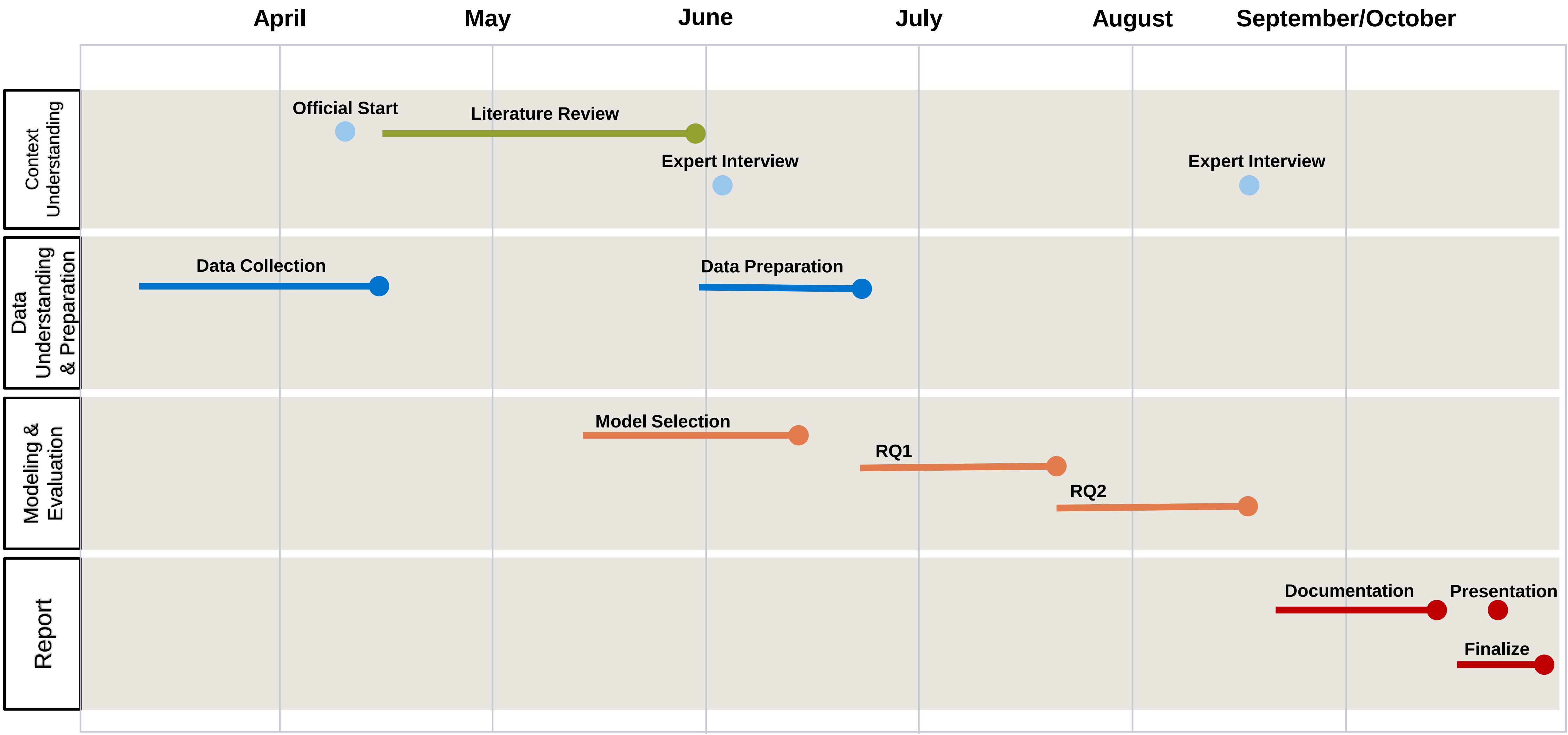
Preprocess Data

Modeling

Evaluation

Abbreviations: MCI = Mild Cognitive Impairment, DM = Dementia, HC = Healthy Control, CTP = Cookie Theft Picture

6. Timeline


April
May
June
July
August
September/October
Context Understanding
Official Start
Literature Review
Expert Interview
Expert Interview
Data Understanding & Preparation
Data Collection
Data Preparation
Modeling & Evaluation
Model Selection
RQ1
RQ2
Report
Documentation
Presentation
Finalize

7.1 Sources

Literature:

[1] WHO. "Dementia Fact Sheet." https://www.who.int/news-room/fact-sheets/detail/dementia (accessed 21 April, 2025).

[2] A. Wimo et al., "The worldwide costs of dementia in 2019," Alzheimer's & Dementia, vol. 19, no. 7, pp. 2865-2873, 2023, doi: https://doi.org/10.1002/alz.12901.

[3] Silva, C. d. M. G. Loures, L. C. V. Alves, L. C. de Souza, K. B. G. Borges, and M. d. G. Carvalho, "Alzheimer’s disease: risk factors and potentially protective measures," Journal of Biomedical Science, vol. 26, no. 1, p. 33, 2019/05/09 2019, doi: 10.1186/s12929-019-0524-y.

[4] Sebis Chair, “Sebis Workshop – 4. Juli 2024,” Technical University of Munich, 2024. [Online]. Available: https://wwwmatthes.in.tum.de/pages/z5zeisq60d7t/Sebis-Workshop-4.-Juli-2024

[5] A. Shakeri, M. Farmanbar, and K. Balog, "MultiConAD: A Unified Multilingual Conversational Dataset for Early Alzheimer's Detection," *arXiv preprint arXiv:2502.19208,* 2025.

[6] B. MacWhinney, “DementiaBank,” TalkBank, Carnegie Mellon University, 2024. [Online]. Available: https://talkbank.org/dementia/

[7] H. Linus, "Exploring key areas of cognitive function: Memory, attention, & more," 2023/01/31 2023. [Online]. Available: https://linushealth.com/blog/exploring-key-areas-of-cognitive-function.

[8] C. De Looze *et al.*, "Cognitive and Structural Correlates of Conversational Speech Timing in Mild Cognitive Impairment and Mild-to-Moderate Alzheimer’s Disease: Relevance for Early Detection Approaches," (in English), *Frontiers in Aging Neuroscience,* Original Research vol. Volume 13 - 2021, 2021-April-27 2021, doi: 10.3389/fnagi.2021.637404.

[9] V. Taler and N. A. Phillips, "Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review," (in eng), *J Clin Exp Neuropsychol,* vol. 30, no. 5, pp. 501-56, Jul 2008, doi: 10.1080/13803390701550128.

[10] L. Toth *et al.*, "A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech," (in eng), *Curr Alzheimer Res,* vol. 15, no. 2, pp. 130-138, 2018, doi: 10.2174/1567205014666171121114930.

[11] K. D. Mueller, B. Hermann, J. Mecollari, and L. S. Turkstra, "Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks," (in eng), *J Clin Exp Neuropsychol,* vol. 40, no. 9, pp. 917-939, Nov 2018, doi: 10.1080/13803395.2018.1446513.

[12] K. C. Fraser, J. A. Meltzer, F. Rudzicz, and P. Garrard, "Linguistic Features Identify Alzheimer’s Disease in Narrative Speech," *Journal of Alzheimer’s Disease,* vol. 49, no. 2, pp. 407-422, 2016, doi: 10.3233/jad-150520.

[13] S. Reeves *et al.*, "Narrative video scene description task discriminates between levels of cognitive impairment in Alzheimer's disease," (in eng), *Neuropsychology,* vol. 34, no. 4, pp. 437-446, May 2020, doi: 10.1037/neu0000621.

[14] C. Roth, "Boston Diagnostic Aphasia Examination," in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan Eds. New York, NY: Springer New York, 2011, pp. 428-430.

[15] P. S. Ambadi, K. Basche, R. L. Koscik, V. Berisha, J. M. Liss, and K. D. Mueller, "Spatio-Semantic Graphs From Picture Description: Applications to Detection of Cognitive Impairment," (in eng), *Front Neurol,* vol. 12, p. 795374, 2021, doi: 10.3389/fneur.2021.795374.

[16] L. Cummings, "Describing the Cookie Theft picture: Sources of breakdown in Alzheimer's dementia," *Pragmatics and Society,* vol. 10, pp. 151-174, 03/28 2019, doi: 10.1075/ps.17011.cum.

[17] H. Chertkow, "Mild cognitive impairment," Current Opinion in Neurology, vol. 15, no. 4, pp. 401–407, 2002. [Online]. Available: https://journals.lww.com/co-neurology/abstract/2002/08000/mild_cognitive_impairment.1.aspx

[18] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, et al., "Mild cognitive impairment," The Lancet, vol. 367, pp. 1262–1270, 2006. [Online]. Available: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(06)68542-5/abstract

[19] D. Knopman, B. Boeve, and R. C. Petersen, "Essentials of the proper diagnoses of mild cognitive impairment, dementia, and major subtypes of dementia," Mayo Clinic Proceedings, vol. 78, no. 10, pp. 1290–1308, 2003. [Online]. Available: https://www.mayoclinicproceedings.org/article/S0025-6196(11)62851-6/abstract

[20] D. Knopman and R. C. Petersen, "Mild cognitive impairment and mild dementia: a clinical perspective," Mayo Clinic Proceedings, vol. 89, no. 10, pp. 1452–1459, 2014. [Online]. Available: https://www.mayoclinicproceedings.org/article/S0025-6196(14)00622-3/fulltext

[21] J. Morris, "Mild cognitive impairment and preclinical Alzheimer's disease," Geriatrics, Suppl, pp. 9–14, 2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16025770/

[22] V. Diaz and G. H. Rodríguez, "Machine learning for detection of cognitive impairment," 2022.

7.2 Sources

Literature:

[23] A. P. Fard, M. H. Mahoor, M. Alsuhaibani, and H. H. Dodge, "Linguistic-based mild cognitive impairment detection using informative loss," *Computers in Biology and Medicine,* vol. 176, p. 108606, 2024.

[24] R. Shankar, A. Bundele, and A. Mukhopadhyay, "A Systematic Review of Natural Language Processing Techniques for Early Detection of Cognitive Impairment," *Mayo Clinic Proceedings: Digital Health,* p. 100205, 2025.

[25] M. Alsuhaibani, A. P. Fard, J. Sun, F. F. Poor, P. S. Pressman, and M. H. Mahoor, "A Review of Deep Learning Approaches for Non-Invasive Cognitive Impairment Detection," *arXiv preprint arXiv:2410.19898,* 2024.

[26] M. Niemelä, M. von Bonsdorff, S. Äyrämö, and T. Kärkkäinen, "Dementia Classification Using Acoustic Speech and Feature Selection," *arXiv preprint arXiv:2502.03484,* 2025.

[27] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, "Mild cognitive impairment represents early-stage Alzheimer disease," Archives of Neurology, vol. 58, no. 3, pp. 397–405, 2001. [Online]. Available: https://jamanetwork.com/journals/jamaneurology/fullarticle/778838

[28] E. Burke, J. Gunstad, and P. Hamrick, "Global and Local Semantic Coherence of Spontaneous Speech in Persons with Alzheimer's Disease and Healthy Controls," Journal of the International Neuropsychological Society, 2023.

[29] R. Bucks, S. Singh, J. M. Cuerden, and G. K. Wilcock, "Analysis of spontaneous, conversational speech in dementia of Alzheimer type," Aphasiology, vol. 14, pp. 71–91, 2000.

[30] I. Hoffmann, D. Németh, C. D. Dye, M. Pákáski, T. Irinyi, and J. Kálmán, "Temporal parameters of spontaneous speech in Alzheimer's disease," International Journal of Speech-Language Pathology, vol. 12, pp. 29–34, 2010.

[31] M. R. Botezatu, E. Miller, and A. M. Kiselica, "Limited connectedness of spontaneous speech may be a marker of dementia," Frontiers in Aging Neuroscience, 2023.

[32] S. Luz, "Longitudinal Monitoring and Detection of Alzheimer's Type Dementia from Spontaneous Speech Data," in IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), 2017.

[33] H. Lee, S. Suniljit, and Y. S. Ong, "Dynamic Multimodal Sentiment Analysis: Leveraging Cross-Modal Attention for Enabled Classification," arXiv preprint, 2025. [Online]. Available: https://arxiv.org/abs/2501.08085

[34] B. Mocanu and T. Ruxandra, "Active Speaker Recognition using Cross Attention Audio-Video Fusion," in 2022 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 2022, pp. 1–6, doi: 10.1109/EUVIP53989.2022.9922810.

[35] H. Xue and Z. Zhu, "Research and Future Application Analysis of Multimodal Fusion," Highlights in Science, Engineering and Technology – EMIS, vol. 119, 2024.

Figures:

[Figure 1] T. Geere, "IBM AI model predicts onset of Alzheimer's disease by analyzing descriptions of a Cookie Theft," *The Next Web*, Nov. 6, 2020. [Online]. Available: https://thenextweb.com/news/ibm-ai-model-predicts-onset-of-alzheimers-disease-by-analyzing-descriptions-of-a-cookie-theft

[Figure 2] J. Clover, "Alexa Guard Rolls Out to Echo Devices in the U.S.," *MacRumors*, May 14, 2019. [Online]. Available: https://www.macrumors.com/2019/05/14/alexa-guard-rolls-out-echo-devices-us/

[Figure 3] S. Lee, "Getting to know the Mel spectrogram," *Medium*, Sep. 5, 2019. [Online]. Available: https://medium.com/data-science/getting-to-know-the-mel-spectrogram-31bca3e2d9d0

Prof. Dr.

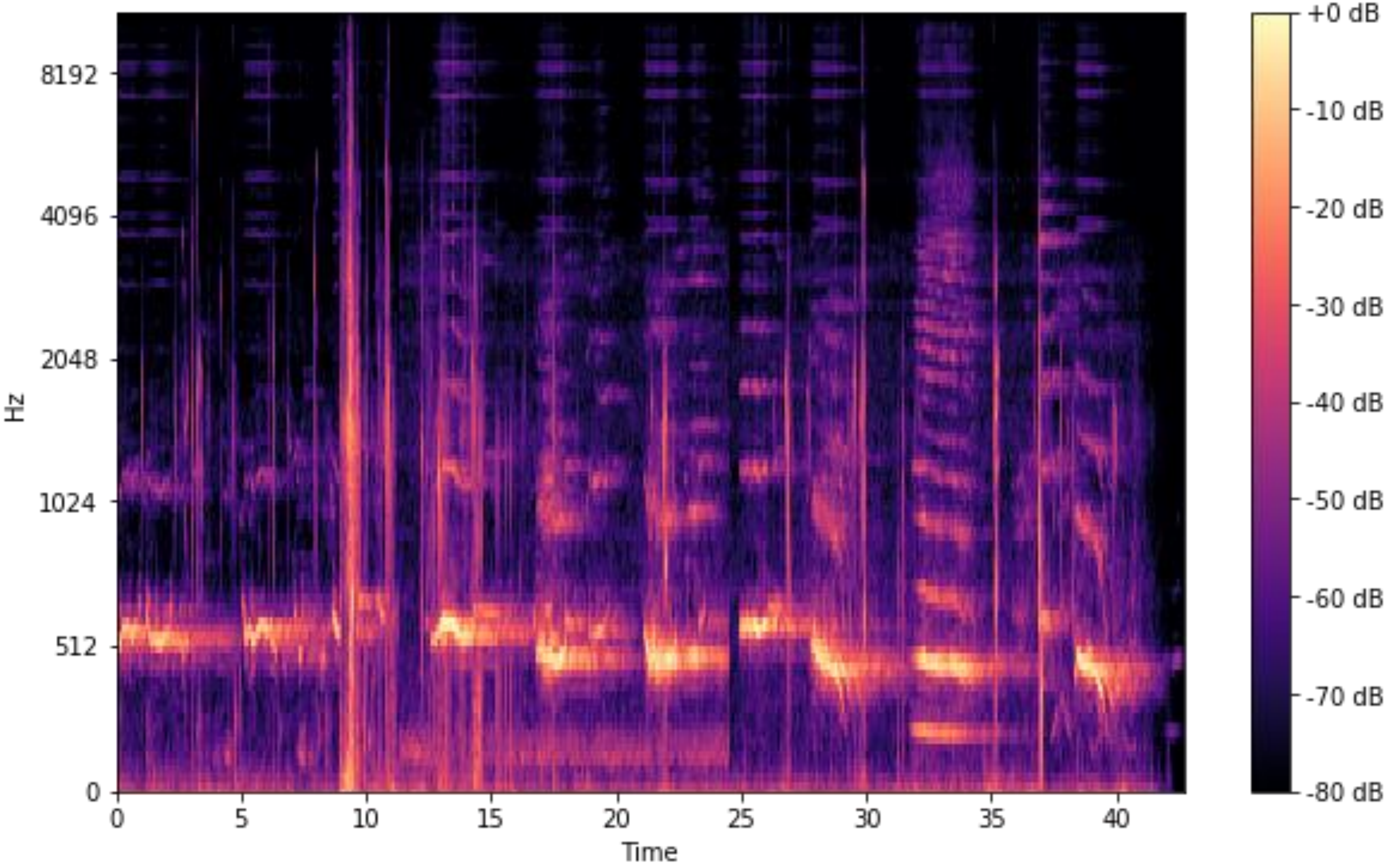
Florian Matthes

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München
17132
Tel +49.89.289.
Fax +49.89.289.17136
matthes@in.tum.de

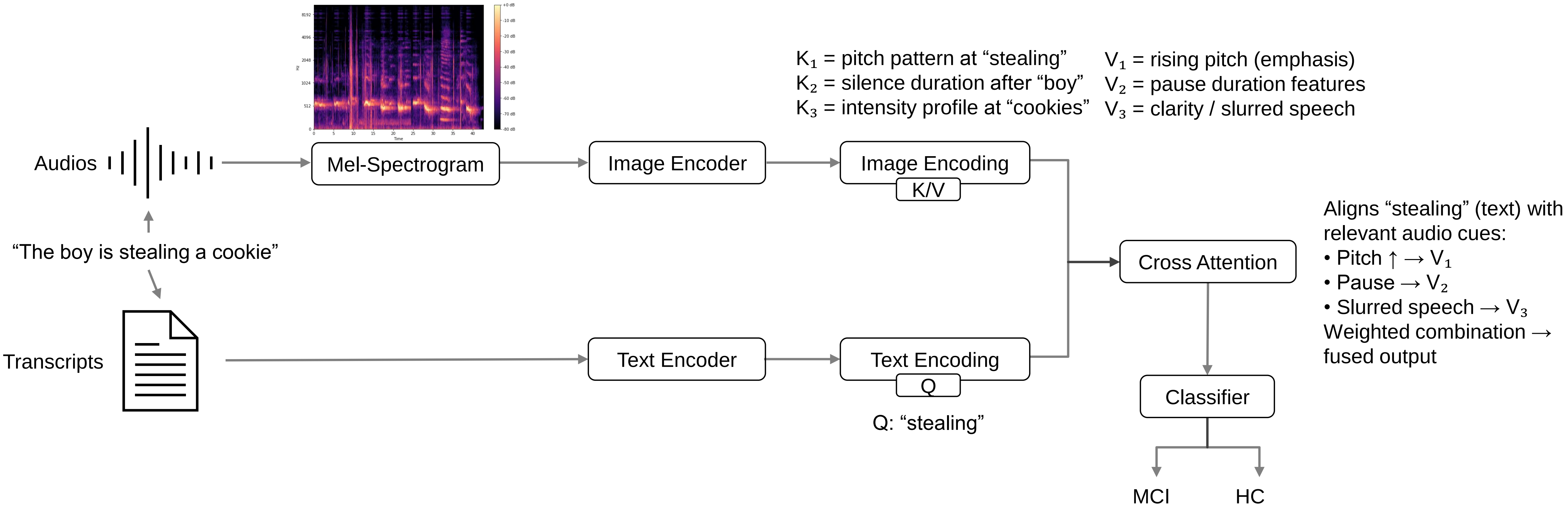
wwwmatthes.in.tum.de

Mel Spectrogram



Source: Figure 3

Cross Attention – Example



Source: Mel-Spectrogram: Figure 3