Friedrich-Alexander-Universität
Department Artificial Intelligence
in Biomedical Engineering

SENSORY
NEURO
ENGINEERING

Transfer Learning for Selective-Attention Decoding in Cochlear-Implant (CI) Users

Neurotechnology Project

at Friedrich-Alexander-Universität Erlangen-Nürnberg
at the Department Artificial Intelligence in Biomedical Engineering (AIBE)
Chair of Sensory Neuroengineering

Principal Supervisor:	Prof. Dr. Tobias Reichenbach
Associate Supervisor:	Constantin Jehn
Author:	Jayana Shah 90439 NURNBERG +49 15758246933 jayana.shah@fau.de 23210894
Submission:	20th February 2025

Abstract

Cochlear implants (CI) have enhanced sound perception for individuals with profound hearing loss, though they still face significant challenges in noisy environments with overlapping speech, known as the "cocktail-party effect". This study investigates the use of Convolutional neural networks (CNNs) and transfer learning to improve Auditory attention decoding (AAD) for CI users. Traditional linear AAD methods are limited in complex auditory environments, while CNN offer superior performance by capturing intricate spatial and temporal Electroencephalography (EEG) features. Our approach adapts CNN pretrained on Normal-hearing (NH) datasets to CI users through partial fine-tuning, addressing the scarcity of labeled CI EEG data and optimizing for the specific neural patterns of CI users. Results demonstrate that this method significantly improves model accuracy and robustness, particularly at a learning rate of 6e-6. By comparing subject-specific and population models, we identify trade-offs in personalization and scalability, highlighting the potential of scalable models for clinical applications. This research contributes to developing effective assistive hearing technologies, enhancing CI users communication in challenging environments.

Contents

Figures

Tables

Abbreviations

AAD	Auditory Attention Decoding
CI	Coclear Implant
CNN	Convolution Neural Network
DNN	Deep Neural Network
EEG	Electroencephalography
FDR	False Discovery Rate
ICA	Independent Component Analysis
NH	Normal Hearing
RNN	Recurrent Neural Network

1 Introduction

1.1 Motivation

The ability of people with profound hearing loss to sense noises and comprehend conversation in quiet settings has been completely transformed with Coclear Implant (CI). However, complex auditory situations with background noise and multiple overlapping voice streams, a phenomenon known as the cocktail party effect, continue to present significant issues for CI users (Bronkhorst, 2000). While individuals with Normal Hearing (NH) can selectively focus on a target speaker, CI users often struggle to distinguish speech from competing auditory streams, leading to reduced speech comprehension.

Addressing this limitation requires advancements in Auditory Attention Decoding (AAD), which seeks to identify the speaker that a listener is focusing on by analyzing neural responses (O'Sullivan et al., 2015). By integrating AAD into CI technology, it may be possible to enhance speech comprehension in noisy environments by amplifying the attended speaker's voice while suppressing background noise. This can enable real-time speech enhancement, personalized hearing assistance, and adaptive noise reduction, ultimately improving the communication experience for CI users.

Traditional AAD methods rely on linear models and statistical correlations between speech stimuli and neural recordings, typically obtained through Electroencephalography (EEG). Although these approaches have provided foundational insights, their performance deteriorates in noisy and overlapping auditory conditions. More recently, Deep Neural Network (DNN)s, particularly Convolution Neural Network (CNN)s, have shown promise in improving AAD performance by capturing complex temporal and spatial patterns in neural data (Puffay et al., 2023).

However, a significant research gap remains in adapting these models for CI users, who exhibit distinct neural responses due to the unique signal processing of cochlear implants. Furthermore, the limited availability of labeled EEG data for CI users presents a challenge for training robust models. This project addresses these limitations by exploring transfer learning techniques to adapt models trained on NH individuals for CI users, with a focus on optimizing CNN architectures for noisy environments.

The practical implications are significant: integrating advanced AAD into CI technology could improve communication in noisy settings, enhancing the quality of life and social integration for CI users.

1.2 Prior Work

The field of AAD has evolved significantly from its initial reliance on classical methods to the adoption of sophisticated neural network approaches. The following sections outline these developments and their relevance to this project.

1.2.1 Classical Approaches

Early AAD methods primarily relied on linear models and statistical methods to establish correlations between EEG signals and auditory stimuli. These approaches laid the groundwork for understanding auditory attention by using techniques such as time-frequency analysis and linear regression to decode attentional focus from EEG signals. Despite their foundational role, these methods are limited by their inability to handle nonlinear interactions in more challenging auditory environments (Emina Alickovic & Ljung, 2019).

1.2.2 Advancements with Neural Networks

The transition to neural network approaches represents a pivotal advancement in AAD. Various neural network architectures, including fully connected networks (Thornton et al., 2023), recurrent long short-term memory models (Monesi et al., 2020), and notably CNN (Accou et al., 2023), have been explored to enhance decoding performance. These models have consistently outperformed linear approaches (see (Puffay et al., 2023) for a comprehensive review).

Each architecture brings unique strengths: Recurrent Neural Network (RNN) are well-suited for modeling temporal dependencies, making them effective in capturing neural data sequences, while CNNs excel in spatial feature extraction, which is essential for decoding neural responses in dynamic listening environments (Puffay et al., 2023). This capacity to model intricate temporal and spatial features has positioned neural networks, especially CNNs, as critical components in advancing AAD techniques.

(Puffay et al., 2023) also explore specific CNN architectures that have been optimized for use in AAD. The Multilayer CNN networks enable the effective extraction of temporal and spatial features from EEG data, improving the model's ability to differentiate attended auditory stimuli from background noise.

Building on these advancements, CNN-based AAD has demonstrated significant improvements in decoding accuracy, enhancing speech comprehension in cocktail party scenarios (Thornton et al., 2023). However, a notable challenge in this research is the

limited availability of labeled training data, which constrains the efficacy and generalization of the models.

1.3 Transfer Learning in AAD

Transfer learning emerges as a dynamic approach to leverage knowledge from one domain or dataset to improve performance in a related but distinct application (Dumoulin, 2023). One fundamental aspect of transfer learning is that the tasks should share similar inputs and exhibit some commonality. This approach is particularly beneficial in data-limited scenarios, such as those involving CI users, where acquiring large-scale labeled datasets is challenging (Weiss et al., 2020).

In the context of AAD, transfer learning enables the adaptation of models trained on large datasets of NH individuals to improve decoding performance for CI users (Weiss et al., 2020). CNN-based architectures benefit from this approach by reusing the lower-layer feature representations, which capture generalizable auditory patterns, while fine tuning the higher layers to adapt to the specific neural responses of CI users (Dumoulin, 2023). This strategy enhances model robustness and improves its ability to generalize to real-world auditory conditions.

By integrating transfer learning into AAD strategies, this project aims to address the existing data limitations and enhance auditory processing for CI users. The objective is to explore how pretrained CNN-based models can be effectively adapted to improve speech comprehension in noisy environments, ultimately contributing to the development of more effective assistive hearing technologies (Jehn et al., 2024).

2 Methodology

2.1 Data Collection and Pre-processing

2.1.1 Data Collection

EEG datasets for both CI users and NH individuals were collected following the procedure by (Jehn et al., 2024). The NH dataset consisted of 28 participants, and the CI dataset included 25 CI users, all of whom were native German speakers and of similar age (refer to Table 1). Both groups performed selective auditory attention tasks in a competing-speaker scenario, focusing on one of two overlapping speech signals.

Metric	**CI Users (N=25)**		**NH Individuals (N=28)**	
	Age (years)	Sex	Age (years)	Sex
Med	56	15F 10M	51	21F 7M
Range	25–81	–	20–72	–
SD	16.7	–	16.22	–

Table 1 Demographic comparison of bilateral CI users and NH individuals. The number of individuals (N), median (Med), range, and standard deviation (SD) are reported for age in years. Sex distribution is presented as the number of females (F) and males (M).

Each participant completed 12 trials, with two simultaneous speech signals (target and distractor) as shown in Table 2. The target (s1 or s2) was alternatively assigned, and participants were instructed to attend to the assigned speaker. The speech signals were presented via loudspeakers positioned at fixed locations as shown in Figure 1.

The auditory stimuli were excerpts from two German audiobooks: “Elbenwald - Blatt von Tuftler” (s1) and “Eine Frau erlebt die Polarnacht” (s2), with s1 slowed to 90% speed for consistency.

Target	**Left Loudspeaker**	**Right Loudspeaker**
s1	s1	s2
s1	s2	s1
s2	s1	s2
s2	s2	s1
..	..	..

Table 2 Competing-speaker scenario for auditory attention tasks. Participants attended to one of two overlapping speech signals (s1 or s2) presented through left and right loudspeakers across 12 trials. The target speaker was alternatively assigned for each trial, simulating real-world multi-speaker environments (Jehn et al., 2024).

EEG signals were recorded using a 32-channel ActiCap system (Brain Products GmbH) at a sampling rate of 1 kHz. A low-pass finite impulse response (FIR) filter with a cutoff

frequency of 280 Hz was applied to the raw EEG to attenuate high-frequency noise. Electrode impedance was maintained below 20 kΩ throughout the recording to ensure signal quality.

To achieve precise synchronization between the EEG and auditory stimuli, onset triggers were used and aligned with stimulus presentation. Any timing discrepancies between the recorded EEG and the audio signal were corrected offline by computing the temporal shift that maximized the Pearson correlation coefficient between the recorded stimulus-evoked audio envelope and its clean reference signal, following the method described in Jehn et al. (2024).

The neural responses to the target and distractor signals were analyzed to study selective auditory attention in both CI users and NH individuals.

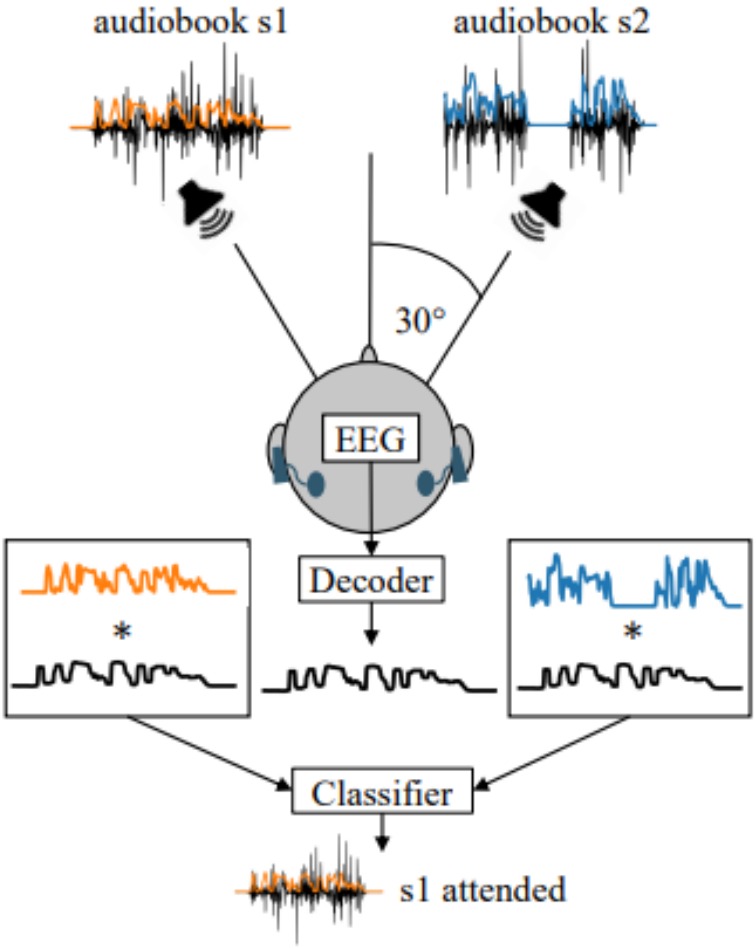


Figure 1 The figure illustrates the experimental setup, with loudspeakers positioned at fixed locations to deliver overlapping speech signals. Participants were instructed to focus on the target speaker while ignoring the distractor (Jehn et al., 2024).

2.1.2 Data Pre-processing

- **Speech Envelope Extraction:** The Hilbert transform was applied to extract speech envelopes, followed by low-pass filtering (56.25 Hz cutoff) and resampling to 125 Hz using MNE-Python version 1.5.1. The onset envelope was also calculated as the rectified first derivative of the speech envelope.
- **EEG Filtering and Resampling:** Spline interpolation was applied to estimate missing EEG channels from neighboring electrodes. EEG signals were low-pass filtered at 36 Hz cutoff and high-pass filtered at 0.5 Hz cutoff, then resampled to 125 Hz to remove noise and slow drifts.
- **Artifact Removal and Standardization:** Independent Component Analysis (ICA) was performed using FastICA algorithm to remove artifacts (e.g., eye movements, cardiac activity, CI stimulation), and EEG channels were standardized to zero mean and unit variance for each trial.

These steps ensured the consistency and comparability of EEG signals across participants and conditions for both CI and NH, allowing for reliable subsequent analysis.

2.2 Model Architecture

2.2.1 CNN Architecture

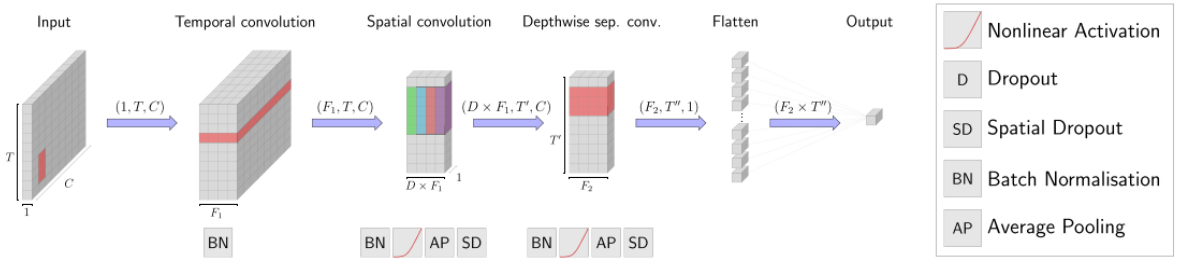


Figure 2 Architecture of the EEGNet model for AAD. The model consists of three main layers: (1) Temporal Convolution for time-domain feature extraction, (2) Spatial Convolution for learning spatial patterns across EEG channels, and (3) Depthwise Separable Convolution for reducing computational complexity while maintaining feature relevance. The architecture is optimized for real-time EEG signal processing (Thornton et al., 2023).

The CNN architecture used in this project is based on EEGNet (Lawhern et al., 2018; Thornton et al., 2023), following the approach outlined in (Jehn et al., 2024). EEGNet

is a lightweight CNN model specifically designed for EEG signal processing, balancing performance with computational efficiency.

The model consists of three main convolutional layers:

- **Temporal Convolution Layer**: Captures time-domain features using 1D convolutional filters.
- **Spatial Convolution Layer**: Learns spatial patterns across EEG channels using grouped convolutions.
- **Depthwise Separable Convolution**: Reduces computational complexity while maintaining meaningful features.

The output of the last convolutional layer is flattened and mapped to the speech feature space using a fully connected linear layer. The grouped convolution approach significantly reduces the number of learnable parameters.

2.2.2 Subject Specific Models (K-fold cross validation)

K-fold cross-validation was applied on both CI users and NH individuals using a leave-one-trial-out approach. Each subject participated in 12 trials. The data was split such that for each subject, one trial was held out as the test set, one as validation set while the remaining 10 trials were used for training. This ensured that every trial was evaluated as the test set at least once, allowing for reliable performance evaluation across different trials within each subject.

The CNN model was optimized by minimizing the negative correlation between the reconstructed and target envelopes. The NAdam optimizer, which incorporates Nesterov momentum, was used to update the model parameters. Hyperparameters such as learning rate (5e-5), batch size (256), dropout rate (0.35), and weight decay (1e-8) were selected based on training and validation performance. The model was trained for up to 10 epochs, with early stopping to prevent overfitting.

2.2.3 Subject Independent Models (Population Model)

To enhance generalization, data from all subjects was pooled together, and the trials were randomly split into training, validation, and test sets. Each subject participated in 12 trials, and one trial was used for validation, another for testing, and the remaining 10 were used for training. This random split allowed the model to capture shared neural patterns across subjects while maintaining balanced evaluation across trials, facilitating

robust model performance at a population level. This was performed on both domains of CI users and NH individuals. Similar to K-Fold Cross-Validation a CNN architecture optimized with the NAdam optimizer was used, with hyperparameters kept consistent.

2.2.4 Comparison of Subject Specific and Subject Independent Models

Both Subject Specific and Subject-Independent approaches were used to balance personalized performance and generalization.

- **K-Fold Cross-Validation (Subject Specific)** This approach ensures that the model is trained and validated on diverse trials within a participant, maximizing robustness for personalized decoding. However, it may struggle to generalize to unseen individuals due to inter-subject variability in EEG signals.
- **Population Model (Subject-Independent)** This method enables the model to learn shared neural representations across individuals, improving generalization to new subjects. However, inter-subject variability and noise can introduce challenges in achieving high decoding accuracy.

2.3 Transfer Learning

Transfer learning is crucial in adapting models trained on NH data for CI users, especially considering the limited availability of labeled CI data. By leveraging knowledge from NH datasets, transfer learning enhances model performance for AAD tasks in CI users.

2.3.1 Fine Tuning with Partial Layers

To enhance generalization to CI data and minimize overfitting, a partial fine tuning approach was used. The key steps included:

- **NH Population Model to CI Data:** A model trained on NH data was used as the base model. This Subject-Independent population model was then adapted to CI data.
- **Freezing the First Two Convolution Layers:** The first two convolutional layers of the NH Population model were frozen, allowing the deeper layers to fine-tune to the specific patterns of CI users.
- **Reduced Learning Rate:** A reduced learning rate, ranging from 1×10^{-6} to 9×10^{-6}, was applied to prevent drastic updates to the pre-learned features

and minimize the risk of overfitting to CI data. The best performance was observed at a learning rate of 6×10^{-6}.

- **Fine Tuning for Individual CI Subjects:** The NH Population model, adapted to CI data, was used as the base model and fine-tuned for each individual CI subject using K-fold cross-validation for Subject Specific adaptation.

As shown in Table 3, this method balances leveraging pre-trained features and adapting to CI-specific data, making it the most suitable for CI EEG adaptation, with less risk of overfitting compared to Full network Fine tuning or Multi-task learning.

Technique	**Advantages**	**Challenges**	**Summary**
Feature Extraction	Fast training, Minimal data required, Effective when datasets are similar.	Limited adaptability, does not address domain differences.	Efficient but insufficient for CI EEG due to domain mismatch.
Knowledge Distillation	Transfers knowledge with limited data, Improves model efficiency.	Fails to handle domain shift, potential accuracy loss.	Not suitable for CI data due to performance trade-offs.
Adapter Layers	Reduces overfitting, Efficient fine tuning on small datasets.	Less adaptable to complex tasks, may not capture CI features.	Efficient but lacks flexibility for domain adaptation.
Multi-Task Learning	Learns shared representations, Improves generalization.	Requires large labeled datasets, Complex model design.	Risky due to domain shift and CI dataset limitations.
Fine Tuning (Partial Layers)	Adapts pre-trained features, Balances efficiency and specificity.	Careful layer selection required, Overfitting risk.	Potentially well suited for CI EEG adaptation with minimal overfitting.
Fine Tuning (Full Network)	Maximizes adaptation to CI-specific EEG.	Needs large datasets, High risk of overfitting.	Too data-intensive for CI users, Impractical.
Domain Adaptation	Addresses domain mismatch, Generalizes across datasets.	Computationally complex, requires tuning.	Strong candidate for CI EEG due to direct domain alignment.

Table 3 Comparison of transfer learning techniques for adapting models from NH to CI Users. The table summarizes the advantages, challenges, and suitability of various transfer learning approaches, highlighting partial fine tuning and domain adaptation as the potentially well suited strategies for addressing domain mismatch and data scarcity in CI EEG data.

2.3.2 Potential for Domain Adaptation

While partial fine tuning effectively adapts the model to CI data, domain adaptation offers further improvement potential. By addressing domain mismatch directly, domain adaptation could improve model robustness by aligning feature distributions between NH and CI data. Techniques such as adversarial training and feature alignment could enhance model generalization, particularly in complex auditory environments.

3 Results

To evaluate the effectiveness of the proposed transfer learning approach for AAD in CI users, I compared the Subject Specific (K-fold cross-validation) and Subject-Independent (Population) models. The models were assessed based on their classification accuracy across different decision windows, with statistical significance evaluated using unpaired "t-tests" and corrected for multiple comparisons using the False Discovery Rate (FDR) procedure.

3.1 Subject Specific Models (K-fold cross validation)

3.1.1 Unpaired "t-test" with FDR

An unpaired "t-test" was performed for each decision window (ranging from 2s to 60s) to evaluate the significance of the classification accuracy differences between NH and CI users. The results are presented in Table 4, where the t-statistics indicate the magnitude of these differences. Larger negative values provide stronger evidence against the null hypothesis (which suggests no significant difference between groups).

Decision Window (s)	T-Statistic	P-Values	Corrected P-Values (FDR)	Degrees of Freedom
2	-4.7251	0.0000	0.0000	51
5	-5.0966	0.0000	0.0000	51
10	-4.6338	0.0000	0.0000	51
20	-4.3321	0.0001	0.0001	51
30	-5.1576	0.0000	0.0000	51
60	-4.7133	0.0000	0.0000	51

Table 4 FDR-Corrected Unpaired "t-test" results comparing classification accuracy between NH and CI users across decision windows. The corrected p-values are still highly significant, confirming that the transfer learning approach enhances performance for both groups.

However, since multiple comparisons were made across different decision windows, the likelihood of encountering false positives increases. To control for this, the p-values were corrected using the FDR procedure. The FDR correction ensures that the proportion of false discoveries (incorrectly rejecting the null hypothesis) is kept below a specified threshold, thus minimizing the risk of drawing invalid conclusions from multiple tests.

The corrected p-values, after FDR adjustment, were found to be statistically significant, with all p-values remaining below the threshold of 0.05 (ranging from 0.0000 to 0.0001). This confirms that the observed differences in classification accuracy between NH and

CI users are unlikely to be due to chance, and the transfer learning approach improves the performance of CI users across all decision windows tested.

The t-statistics, ranging from -4.33 to -5.16, further support these findings, with the negative values indicating that CI users have lower accuracy than NH users, but the transfer learning approach significantly narrows this gap.

3.1.2 Impact of Decision Window Length

The classification accuracies for both CI and NH users across different decision windows for the Subject Specific (K-fold cross-validation) model are presented in Table 5. The results demonstrate that longer decision windows lead to improved classification accuracy for both groups.

For CI users, the accuracy improved from 55.8% at 2s to 71.3% at 60s, while for NH users, accuracy increased from 60.9% to 87.0% 5. These findings suggest that longer decision windows allow the model to capture more stable neural patterns associated with auditory attention.

Decision Window (s)	CI Users Accuracy (%)	NH Users Accuracy (%)
2	55.80	60.98
5	58.68	67.08
10	62.12	71.97
20	65.82	77.55
30	67.65	81.63
60	71.33	87.05

Table 5 Classification accuracy (%) of CI and NH users across different decision windows for the Subject Specific Model.

The plots of classification accuracies for both groups are shown in Figures 3 and 4, providing a visual representation of the improvement in accuracy with longer decision windows.

3.2 Subject Independent Models (Population Model)

3.2.1 Unpaired "t-test" with FDR

For the Population model, an unpaired "t-test" was also conducted to assess the classification accuracy differences between CI and NH users. The results, corrected for multiple comparisons using the FDR procedure, are shown in Table 6.

Decision Window (s)	T-Statistic	P-Values	Corrected P-Values (FDR)	Degrees of Freedom
2	-2.8123	0.0070	0.0169	51
5	-2.9454	0.0049	0.0169	51
10	-2.7356	0.0086	0.0169	51
20	-2.6316	0.0112	0.0169	51
30	-2.1097	0.0399	0.0479	51
60	-1.7769	0.0818	0.0818	51

Table 6 FDR-Corrected Unpaired "t-test" results comparing classification accuracy between NH and CI users across decision windows for the Population Model.

The results show that for decision windows from 2s to 30s, the classification accuracy differences between CI and NH users were statistically significant, with corrected "p-values" below 0.05. However, for the 60s decision window, with the "p-values" of 0.0818 did not reach statistical significance, indicating the gap between groups narrows as the decision window lengthens

3.2.2 Impact of Decision Window Length

The classification accuracies for both CI and NH users across different decision windows for the Population model are summarized in Table 7. For CI users, the accuracy improved from 52.64% at 2s to 65.67% at 60s, while for NH users, accuracy increased from 55.30% to 72.62%. These results are consistent with the trends observed in the Subject Specific model, suggesting that longer decision windows provide more reliable features for classification. The plots of classification accuracies for both groups are shown in Figures 3 and 4 for the Subject Independent Model.

Decision Window (s)	CI Users Accuracy (%)	NH Users Accuracy (%)
2	52.64	55.30
5	54.83	58.99
10	56.43	61.47
20	58.83	65.39
30	61.83	67.57
60	65.67	72.62

Table 7 Classification accuracy (%) of CI and NH users across different decision windows for the Population Model.

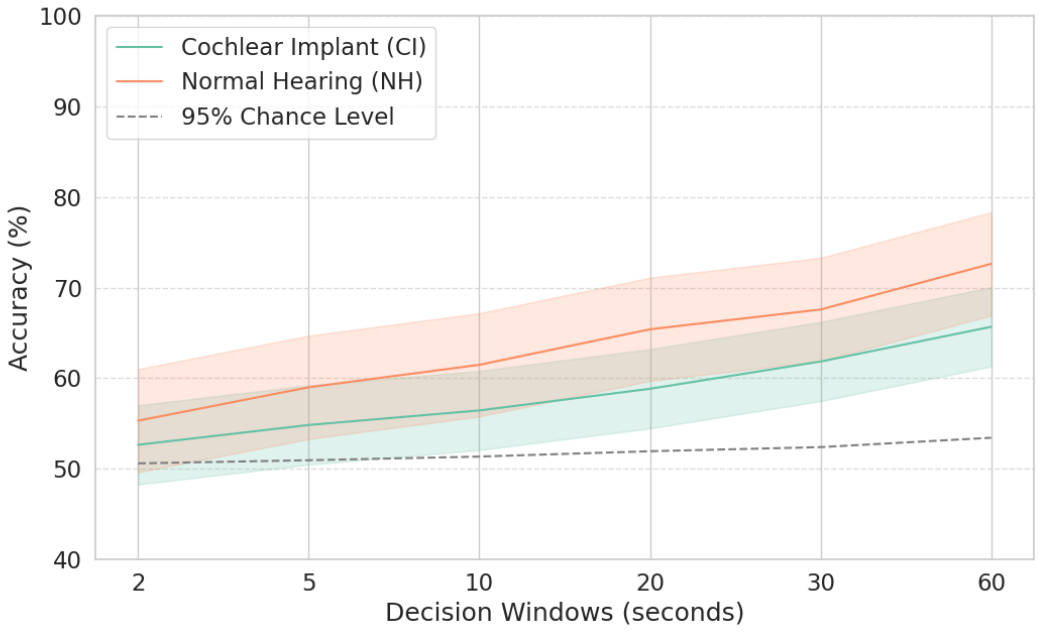
3.3 Comparison of Subject Specific and Subject Independent Models

This project compared two models: the **Subject Specific Model (K-fold Cross-Validation)** and the **Subject-Independent Model (Population Model)**. Key findings include:

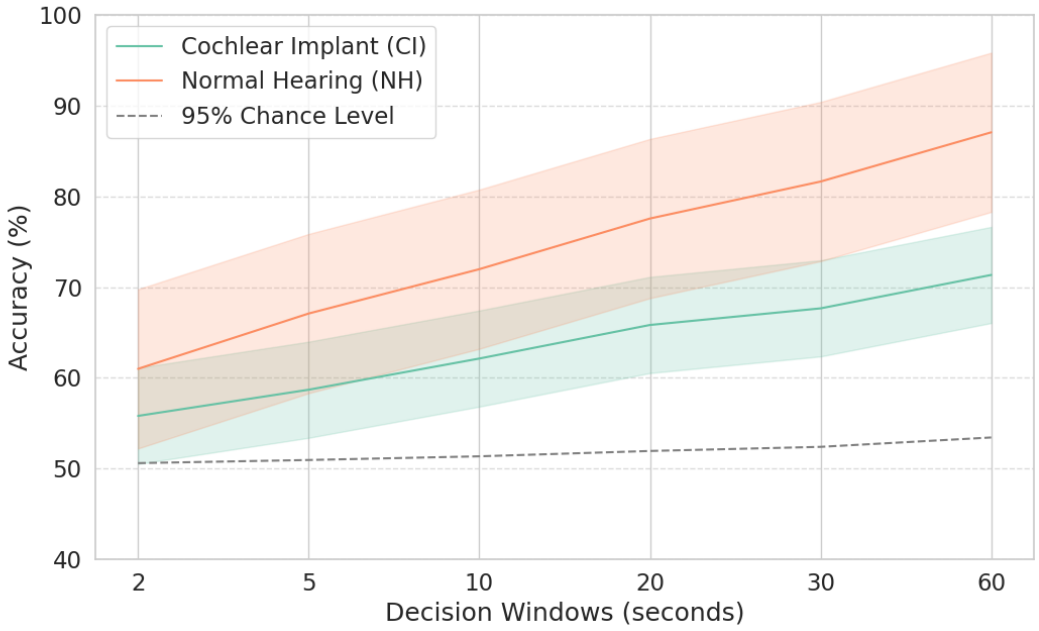
- **Personalization vs. Generalization:** The Subject Specific model achieved higher accuracy for known subjects (e.g., 71.33% at 60s) but struggled to generalize to new individuals. In contrast, the Subject-Independent model offered better generalization (e.g., 64.10% at 60s) at the cost of slightly lower personalized performance.
- **Real-World Applicability:** The Subject-Independent model is more scalable and practical for scenarios where individual data may not be consistently available, such as in clinical settings or consumer devices.
- **Computational Complexity:**The Subject Specific model required more computational resources for training on individual data, while the Subject-Independent model was more efficient by pooling data across subjects.

Figures 3 and 4 showcase the performance differences, highlighting the trade-offs between personalized accuracy and generalization ability.

In the line plots presented in Figures 3, a 95% chance level is plotted as a reference, representing the baseline or expected accuracy level achieved by random chance. It refers to the accuracy you would expect if the classifier were guessing randomly, based on the number of classes.

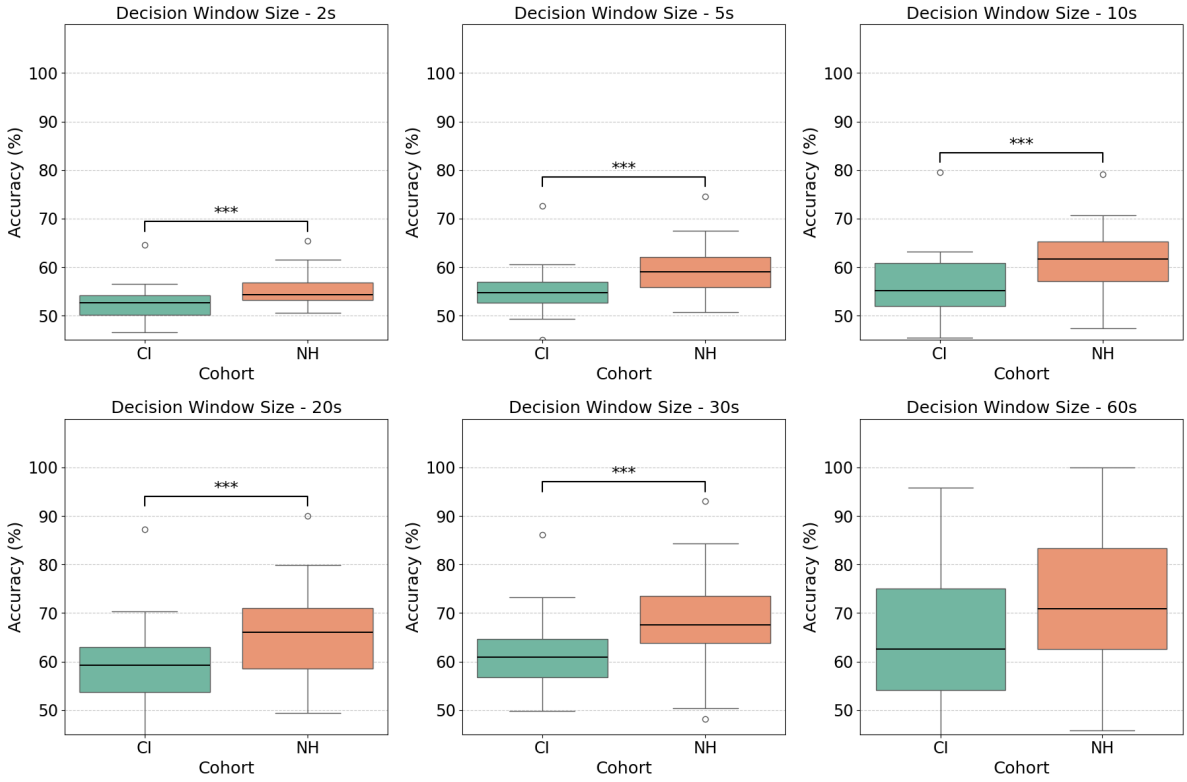


Subject Independent Model (Population Model)

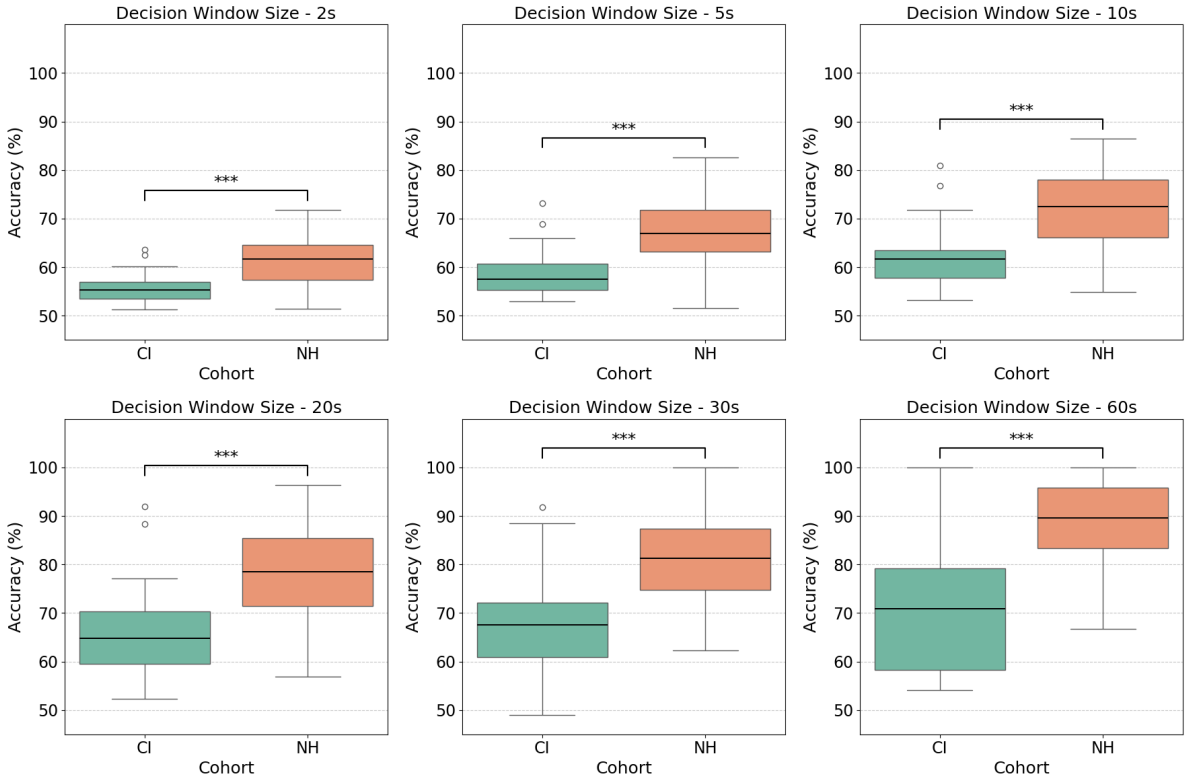


Subject Specific Model (K-fold Cross Validation)

Figure 3 Line Plot comparing classification accuracy across decision windows for Subject Specific and Subject-Independent Models. Shaded error bands represent the standard deviation, highlighting the trade-off between personalization and generalization.



Subject Independent Model (Population Model)



Subject Specific Model (K-fold Cross Validation)

Figure 4 Box Plot comparing classification accuracy across decision windows for Subject Specific and Subject-Independent Models. The box plots illustrate the distribution of accuracy values, emphasizing the variability in model performance.

3.4 Transfer Learning

3.4.1 Fine Tuning with Partial Layers

Building on the previous section's results, transfer learning from NH datasets to CI users was explored as a means to improve decoding accuracy. The fine tuning approach involved adapting pre-trained models on NH data to better fit the characteristics of CI users, by freezing the initial layers and fine tuning the deeper layers using various learning rates. The results, presented in Figure 5 and Table 8, highlight the performance of the model across different decision windows.

Model / Learning Rate	2s	5s	10s	20s	30s	60s
Subject Specific CI Model	55.7987	58.6824	62.1192	65.8187	67.65	71.3333
Subject-Independent CI Model	52.9019	54.9177	56.6108	59.3488	60.9178	64.1007
Fine Tuning (LR 1e-5)	54.1171	58.1411	60.7744	64.1941	68.3611	72.0
Fine Tuning (LR 3e-5)	52.8334	55.0144	57.3099	59.7484	64.5556	67.5
Fine Tuning (LR 1e-6)	54.5136	57.4023	60.2335	63.1163	66.9944	70.0
Fine Tuning (LR 5e-6)	54.6655	58.1791	61.3697	64.7425	69.6889	70.6667
Fine Tuning (LR 6e-6)	55.0799	58.6366	61.0478	65.0595	69.2889	74.3333
Fine Tuning (LR 9e-6)	54.2862	57.4343	59.5133	64.3623	68.3444	70.5

Table 8 Performance comparison of models across decision windows. Fine tuning with a learning rate of 6e-6 achieved the highest accuracy

The results show that transfer learning through fine tuning produced varied improvements across decision windows. The fine tuning with a learning rate of 6e-6 slightly outperformed the baseline models, especially in the longer decision windows. The most notable improvements were observed in the 30-second (69.29% vs. 67.65% for the Subject Specific model) and 60-second decision windows (74.33% vs. 71.33% for the Subject Specific model), where the fine-tuned model showed significant gains in accuracy.

To validate the significance in accuracy, the models were evaluated on individual subjects from the test set. This evaluation helped ensure that the observed gains in performance were consistent across subjects and decision windows, rather than due to random variation or overfitting to specific data points.

For shorter decision windows (2s to 20s), the improvements from fine tuning were more modest. For instance, in the 2-second window, the fine-tuned model achieved an accuracy of 55.08%, which is slightly better than the Subject Specific CI model (55.80%), but the difference became less pronounced in the shorter windows. Similarly, in the 5-second and 10-second windows, the fine-tuned model showed marginal improve-

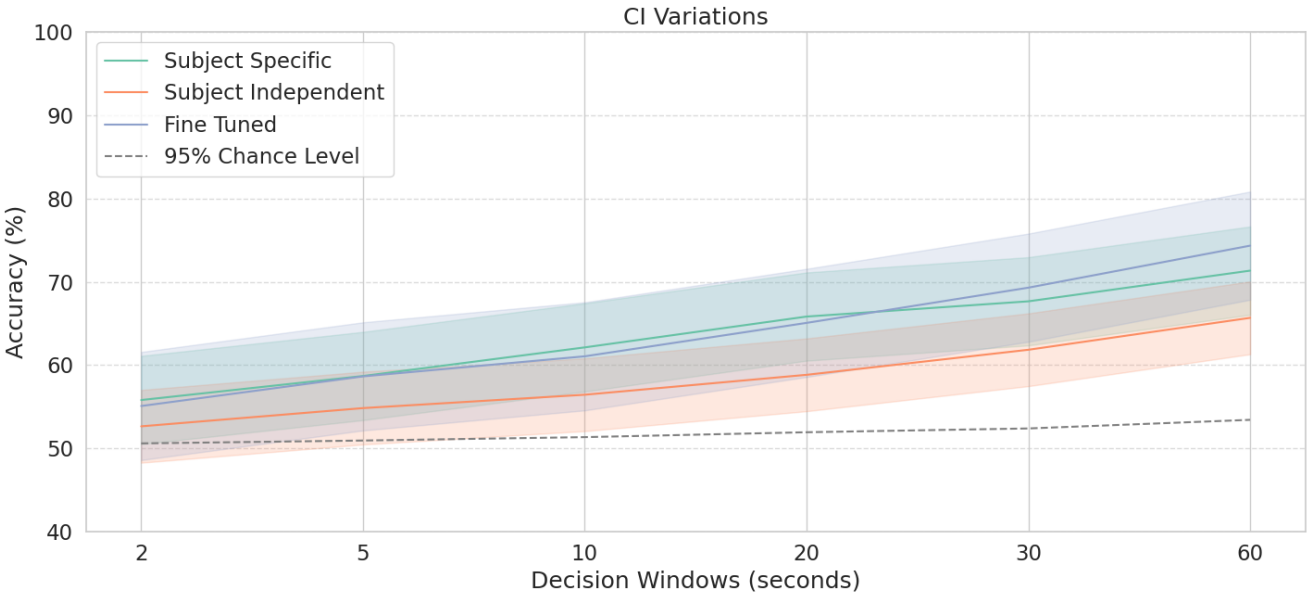


Figure 5 The figure illustrates the adaptation of a pre-trained NH model to CI data by freezing the initial layers and fine tuning the deeper layers with reduced learning rate.

ments over the Subject Specific model (58.64% vs. 58.68% and 61.05% vs. 62.12%, respectively).

Interestingly, the fine tuning approach with a learning rate of 6e-6 led to more consistent improvements across all decision windows compared to the other learning rates tested (such as 1e-5 and 3e-5). While fine tuning with other learning rates also provided some improvement, particularly in the 30s and 60s windows, it did not match the performance achieved with a learning rate of 6e-6.

These findings suggest that transfer learning, especially with a carefully selected learning rate, can be a powerful tool for improving decoding accuracy in CI users. While the improvements are more pronounced in longer decision windows, the overall results highlight the value of fine tuning pre-trained models on NH data to adapt them effectively to the specific characteristics of CI users. This approach can be particularly beneficial when labeled data for CI users is limited, as it allows leveraging the knowledge learned from NH data to boost model performance in challenging scenarios.

4 Discussion

The project has explored the potential of enhancing AAD for CI users through the integration of CNN and transfer learning techniques. This investigation builds on prior research that highlights the challenges faced by CI users in noisy environments, specifically the cocktail party effect, and aims to bridge the performance gap between NH individuals and CI users.

4.1 Key Findings

The results demonstrate that CNN-based AAD significantly enhances decoding accuracy in complex auditory environments. The adoption of neural networks, particularly CNNs, enables the extraction of detailed temporal and spatial patterns from EEG data. This advancement provides a marked improvement over traditional linear models, aligning with previous studies that emphasize the importance of neural network architectures in processing nonlinear neural interactions (Puffay et al., 2023; Thornton et al., 2023).

A major contribution of this project is the application of transfer learning to adapt pretrained models to NH data for CI users. The data scarcity challenge inherent in CI datasets is a known impediment to model training. Our use of transfer learning addresses this limitation by effectively leveraging extensive NH datasets to pretrain models, which are then fine-tuned to the specificities of CI user data. (Dumoulin, 2023; Weiss et al., 2020).

4.2 Comparison of Model Approaches

Comparison between Subject Specific and Subject-Independent models revealed critical insights regarding their respective strengths. While the Subject Specific model offered personalized adaptation and higher accuracy for individual CI users, the Subject-Independent approach demonstrated superior generalization across users, a crucial factor for wider clinical deployment and adaptive hearing solutions. This trade-off between personalization and scalability is an essential consideration for future model designs.

Improvements were only made for longer decision windows, indicating that the model's performance was highly sensitive to the duration of the decision-making process. The Subject Specific model, in particular, benefited more from longer decision windows, while shorter windows did not capture the necessary auditory cues as effectively. This

suggests that refining the decision window length could play a significant role in further enhancing model performance, particularly for the Subject-Independent approach, which might be more constrained by shorter windows.

4.3 Impact of Transfer Learning Techniques

The strategic implementation of transfer learning was key in improving model adaptability. The partial fine tuning approach, involving the freezing of initial layers and fine tuning of deeper ones, proved effective. By preserving the robust, generalizable features extracted from NH data and adapting the later layers to accommodate CI-specific nuances, this methodology resulted in slight accuracy gains at a learning rate of 6e-6.

This finding underscores the potential of transfer learning to transform assistive hearing technologies, facilitating better generalization across user populations. Moreover, our exploration of various learning rates further emphasizes the delicate balance required to optimize fine tuning, minimizing overfitting.

4.4 Limitations and Future Directions

Despite the promising results, limitations persist. The intrinsic variability of EEG signals among individuals presents challenges in achieving a universally applicable model. Moreover, the project's reliance on predefined learning rates and model parameters may not generalize across different datasets or auditory scenarios.

Future work should focus on expanding the dataset diversity to include more varied CI user populations and auditory conditions. Exploration of alternative neural architectures and dynamic adjustment of learning parameters may also enhance model performance. Additionally, investigating domain adaptation techniques could further address domain mismatches, potentially elevating the robustness of AAD systems across broader contexts.

5 Conclusion

This project explored the application of transfer learning to improve AAD for CI users in noisy environments. By leveraging pre-trained models from NH individuals and fine tuning them on CI data, we demonstrated improvements in decoding accuracy, particularly with longer decision windows. The results highlight the potential of transfer learning to bridge the gap between NH and CI datasets, addressing the challenge of limited labeled data for CI users.

The comparison between Subject Specific and Subject-Independent models revealed important trade-offs between personalization and generalization. While Subject Specific models achieved higher accuracy for known subjects, the Subject-Independent model offered better generalization across different individuals, making it more scalable for real-world applications. The partial fine tuning approach proved to be particularly effective, as it allowed the model to adapt to CI-specific patterns without overfitting.

Despite these promising results, there are limitations to this project, including the small sample size of CI users and the reliance on non-real time analysis. Future work should focus on expanding the dataset, exploring advanced domain adaptation techniques, and developing real-time AAD algorithms for integration into CI devices.

In summary, this project contributes to the growing body of research on AAD for CI users, offering a robust and scalable solution to improve speech comprehension in noisy environments. By leveraging transfer learning, we can enhance the performance of assistive hearing technologies, ultimately improving the quality of life for CI users.

References

Accou, B., Vanthornhout, J., Van hamme, H., & Francart, T. (2023). Decoding of the speech envelope from eeg using the vlaai deep neural network. *Scientific Reports*, *13*(1), 812.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, *86*(1), 117–128.

Dumoulin, V. e. a. (2023). Transfer learning. In *Deep learning* (pp. 189–191). Springer. https://link.springer.com/book/10.1007/978-3-031-45468-4

Emina Alickovic, F. G., Thomas Lunner, & Ljung, L. (2019). A tutorial on auditory attention identification methods. *Frontiers in Neuroscience*. https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2019.00153/full

Jehn, C., Kossmann, A., Vavatzanidis, N. K., Hahne, A., & Reichenbach, T. (2024). Cnns improve decoding of selective attention to speech in cochlear implant users [Preprint]. *TechRxiv Preprint*. https://doi.org/10.36227/techrxiv.171710242.23399264/v1

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), 056013. https://doi.org/10.1088/1741-2552/aace8c

Monesi, M. J., Accou, B., Montoya-Martinez, J., Francart, T., & Van hamme, H. (2020). An lstm based architecture to relate speech stimulus to eeg. *IEEE*, 941–945.

O'Sullivan, J. A., et al. (2015). Neural decoding of attentional selection in multi-speaker environments. *Current Biology*.

Puffay, C., Accou, B., Bollens, L., Jalilpour Monesi, M., Vanthornhout, J., Van Hamme, H., & Francart, T. (2023). Relating eeg to continuous speech using deep neural networks: A review [Preprint]. *arXiv preprint arXiv:2302.01736*. https://doi.org/10.48550/arXiv.2302.01736

Thornton, M., Mandic, D., & Reichenbach, T. (2023). Robust decoding of the speech envelope from eeg recordings through deep neural networks. *Journal of Neural Engineering*. https://iopscience.iop.org/article/10.1088/1741-2552/ac7976/meta

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2020). A Survey of Transfer Learning. *IEEE Transactions on Big Data*, *10*(1), 223–249. https://doi.org/10.1109/TBDATA.2020.2994743

Declaration of Academic Integrity

I hereby declare that this project and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly attributed. Where I have quoted from the work of others, the source is always given. I am aware that the project in digital form can be examined for the use of unauthorised aid and in order to determine whether the project as a whole or in parts may amount to plagiarism. I am aware that a false assurance fulfils the elements of fraud in accord with § 10 and § 13 ABM-PO/TechFak and will result in the consequences proclaimed there. This paper was not previously presented to another examination board and has not been published.

NURNBERG, 20th February 2025

Jayana Shah