

DeMoAiR: DEEP LEARNING BASED MULTI-MODAL AIRWRITING RECOGNITION FOR SMART WEARABLES

AYUSH TRIPATHI



**DEPARTMENT of ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
MAY 2024**

DeMoAiR: DEEP LEARNING BASED MULTI-MODAL AIRWRITING RECOGNITION FOR SMART WEARABLES

by

AYUSH TRIPATHI

DEPARTMENT of ELECTRICAL ENGINEERING

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

MAY 2024

Dedicated to
My Family



Department of Electrical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi, India-110016

CERTIFICATE

This is to certify that the thesis entitled “**DeMoAiR: Deep Learning based Multi-modal Airwriting Recognition for Smart Wearables**” being submitted by **Mr. Ayush Tripathi** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy** is the record of bonafide research work carried out by him under our supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other University or Institute for the award of any degree or diploma.


Dr. Lalan Kumar
Associate Professor
Department of Electrical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi-110016, India

Dr. Lalan Kumar
Associate Professor,
Department of Electrical Engineering,
Bharti School of Telecommunication,
Yardi School of Artificial Intelligence,
Indian Institute of Technology Delhi,
New Delhi, India

Date:

Place:



Dr. Prathosh A.P.
Assistant Professor,
Department of Electrical Communication Engineering,
Indian Institute of Science (IISc), Bengaluru, India

Dr. Prathosh A.P.
Assistant Professor,
Department of Electrical
Communication Engineering,
Indian Institute of Science (IISc),
Bengaluru, India

Date:

Place:

Acknowledgment

I want to start by expressing profound gratitude towards almighty God for giving me the opportunity to take birth in the human form of life and the wisdom to explore the journey in the research world. My heartfelt appreciation goes to my parents, whose contributions to my life are immeasurable, and no words can truly encapsulate the extent of my gratitude towards them. Special thanks are also due to my younger brother, Adi, whose endearing presence never failed to lift my spirits throughout this journey. I am sincerely grateful to my family for their unwavering support, without which the completion of this dissertation would not have been achievable.

My research journey began when I was in the second year of my B.Tech. program and was offered a summer internship by Prof. S.R. Mahadeva Prasanna at IIT Guwahati. Those two months at IIT Guwahati not only altered my perspective but also instilled in me an enduring motivation to pursue a career in research. During this period, I worked alongside remarkable mentors such as Dr. Akhilesh, Dr. Vikram, and Dr. Sishir, whose insightful discussions inspired me to set my sights on a Ph.D. My research journey further unfolded in the lab of Prof. Visar Berisha at Arizona State University, where I found a welcoming and encouraging environment. I am immensely grateful for the guidance and support provided by Prof. Berisha. Additionally, a significant influence on my decision to pursue a Ph.D. can be attributed to Dr. Sunil Kumar Kopparapu, whom I had the honor of working with during a brief stint at TCS Research and Innovation Labs in Mumbai. He played a crucial role in helping me understand the nuances of research and fostering my aspirations for higher education.

I extend my heartfelt and sincere gratitude to my esteemed supervisors, Prof. Lalan Kumar and Prof. Prathosh A.P., whose unwavering guidance and support have been instrumental in realizing my dream of attaining a Ph.D. Throughout the challenging journey of my doctoral studies, they have consistently stood by me, providing assistance in every possible way during difficult times. I am thankful to Dr. Suriya Prakash Muthukrishnan from the All India Institute of Medical Sciences, New Delhi for providing his insights on the data recording setup. I am also grateful to the SRC members: Prof. Sumantra Dutta Roy, Prof. Seshan Srirangarajan, and Prof. Monika Aggarwal, for their invaluable feedback during the progress presentations. Their

insights have significantly contributed to the refinement of my research work. Furthermore, I express my thanks to the Ministry of Education, Government of India, for awarding the Prime Minister's Research Fellows (PMRF) Scheme for monetarily supporting my Ph.D. The PMRF contingency grant proved to be exceptionally beneficial, facilitating my participation in conferences and the acquisition of essential equipment for my research.

I extend my thanks to all my friends and colleagues who have been a continual presence and a pillar of support throughout the different chapters of my life. A special acknowledgement goes to my mentor, Prof. C.V. Sastri, whose encouragement led me to embark on the journey of pursuing a Ph.D. at a young age. A heartfelt thanks goes to Dr. Arnab Mondal, my senior from IIT Delhi. I am grateful for the privilege of working with him, and his mentorship during times of need has been invaluable. I am deeply thankful to my friends from my B.Tech. days: Brijesh, Shivang, Shubhanshu, Divyansh, Anish, Tushar, Aditya, Saurabh, and Kaushik. Their consistent motivation and companionship played a significant role in keeping me inspired, especially during challenging times. Acknowledgements are also due to my former colleagues from TCS Research and Innovation Labs: Meet, Swapnil Aditya, Upasana, Meghna, Sushrut, Dr. Rupayan, and Dr. Ashish. Our fruitful discussions and casual chit-chats enriched my professional experience. I express heartfelt thanks to my fellow MSP labmates and seniors: Anant, Shivani, Prerna, Deepika, Amita, Ashish ji, and Alka ji for their unwavering support and encouragement. Special appreciation goes to my friends from Saptagiri Hostel, without whom my life during Ph.D. would have been no fun.

I express sincere gratitude to everyone who has played a role, whether direct or indirect, in supporting me throughout this journey. I am thankful for the collective assistance that has shaped my path.

Abstract

The ability to communicate stands out as a fundamental life skill for humans. As digital devices are rapidly evolving, the interaction between humans and these devices is also increasing. Conventionally, users engage with computers through peripheral devices such as keyboards, mouse, or touchscreen. However, the future of technology aspires to eliminate the dependence on these mediating devices, thereby relieving individuals from the burden of carrying additional physical tools. In pursuit of an alternative input method, this thesis explores the field of multi-modal airwriting recognition. The objective is to provide users with an innovative means of input and therefore contribute to the landscape of human-computer interaction. Airwriting is defined as the act of freely writing letters in the air through unrestricted finger movements, constituting a specialized form of gesture recognition. Such a system does not require users to learn a new set of gestures and also allows for a versatile range of words formed by concatenating letters. This broadens the spectrum of interaction capabilities for users. The process of airwriting necessitates coordination between the brain, forearm muscles, and the wrist to execute the desired movements. Consequently, recognition of airwriting may involve analyzing cues from each of these individual aspects, leading to a multi-modal approach. This approach integrates Electroencephalography (EEG) for brain signals, Electromyography (EMG) for muscle signals, and Inertial Measurement Unit (IMU) for wrist movement. Each modality provides a unique perspective on the same information. This thesis explores the feasibility of airwriting recognition utilizing diverse modalities.

The thesis begins with a focus on airwriting recognition using signals recorded from a wrist-worn IMU. A supervised contrastive loss based framework for airwriting recognition (SCLAiR) is presented. The framework employs a 2-stage classification method: in the first stage, supervised contrastive loss is applied, followed by the minimization of cross-entropy loss in the second stage. This approach enhances classification accuracy on both a publicly available dataset and a dataset recorded in the lab. Additionally, a real-time demonstration showcasing airwriting recognition is provided. Subsequently, image representation of time-series data (ImAiR) is also utilized for airwriting recognition. First, the time-series data is encoded to image representation and then deep learning-based models are utilized for identifying the

written alphabets. Different techniques such as Self Similarity Matrix, Gramian Angular Field, and Markov Transition Field, are utilized for image encoding. Various standard model architectures for image classification are then used for classification.

The prospect of airwriting recognition using EMG signals recorded from forearm muscles is subsequently explored. The SurfMyoAiR dataset comprising EMG signals while writing English uppercase alphabets is constructed. Different time-domain features to construct EMG envelopes and different time-frequency image representations are explored to form the input to a deep learning model for airwriting recognition. Several different deep learning architectures are exploited for this task. Furthermore, a multi-loss minimization framework for EMG-based airwriting recognition (TripCEAiR) is proposed. The framework aims at learning a feature embedding vector that minimizes the triplet loss, while simultaneously learning the parameters of a classifier head to recognize corresponding alphabets. The proposed method is validated on the SurfMyoAiR dataset. The effect of different variations of triplet loss, triplet mining strategies, and feature embedding dimension is also comprehensively investigated.

Leveraging EEG signals for airwriting recognition offers a promising alternative input method for Human-Computer Interaction and is explored as a part of this thesis. The NeuroAiR dataset comprising EEG signals recorded while writing English uppercase alphabets is first constructed. Various features are then explored in conjunction with different deep learning models to achieve accurate airwriting recognition. These features include processed EEG data, Independent Component Analysis (ICA) components, source-domain-based scout time series, and spherical and head harmonic decomposition-based features. Furthermore, the impact of different EEG frequency bands on system performance is comprehensively examined. This sets a strong baseline for future advancements and demonstrates the viability and utility of EEG-based airwriting recognition.

सारांश

संचार करने की क्षमता मनुष्य के लिए एक मौलिक जीवन कौशल के रूप में सामने आती है। जैसे-जैसे डिजिटल उपकरण तेजी से विकसित हो रहे हैं, इंसानों और इन उपकरणों के बीच संपर्क भी बढ़ रहा है। परंपरागत रूप से, उपयोगकर्ता कीबोर्ड, माउस या टचस्क्रीन जैसे परिधीय उपकरणों के माध्यम से कंप्यूटर से जुड़ते हैं। हालाँकि, प्रौद्योगिकी का भविष्य इन मध्यस्थ उपकरणों पर निर्भरता को खत्म करने की आकांक्षा रखता है, जिससे व्यक्तियों को अतिरिक्त भौतिक उपकरण ले जाने के बोझ से राहत मिल सके। एक वैकल्पिक इनपुट पद्धति की खोज में, यह थीसिस मल्टी-मॉडल एयरराइटिंग मान्यता के क्षेत्र की पड़ताल करती है। इसका उद्देश्य उपयोगकर्ताओं को इनपुट का एक अभिनव साधन प्रदान करना है और इसलिए मानव-कंप्यूटर संपर्क के परिदृश्य में योगदान करना है। एयरराइटिंग को अप्रतिबंधित उंगली आंदोलनों के माध्यम से हवा में स्वतंत्र रूप से पत्र लिखने के कार्य के रूप में परिभाषित किया गया है, जो हावभाव पहचान का एक विशेष रूप है। इस तरह की प्रणाली के लिए उपयोगकर्ताओं को इशारों का एक नया सेट सीखने की आवश्यकता नहीं होती है और अक्षरों को जोड़कर शब्दों की एक बहुमुखी श्रृंखला बनाने की भी अनुमति मिलती है। यह उपयोगकर्ताओं के लिए इंटरैक्शन क्षमताओं के स्पेक्ट्रम को विस्तृत करता है। वायुलेखन की प्रक्रिया में वांछित गतिविधियों को निष्पादित करने के लिए मस्तिष्क, अग्रबाहु की मांसपेशियों और कलाई के बीच समन्वय की आवश्यकता होती है। नतीजतन, एयरराइटिंग की पहचान में इनमें से प्रत्येक व्यक्तिगत पहलू से संकेतों का विश्लेषण करना शामिल हो सकता है, जिससे मल्टी-मॉडल दृष्टिकोण प्राप्त हो सकता है। यह दृष्टिकोण मस्तिष्क संकेतों के लिए इलेक्ट्रोएन्सेफेलोग्राफी (ईईजी), मांसपेशियों के संकेतों के लिए इलेक्ट्रोमायोग्राफी (ईएमजी) और कलाई की गति के लिए जड़त्व माप इकाई (आईएमयू) को एकीकृत करता है। प्रत्येक पद्धति एक ही जानकारी पर एक अद्वितीय परिप्रेक्ष्य प्रदान करती है। यह थीसिस विविध तौर-तरीकों का उपयोग करके एयरराइटिंग पहचान की व्यवहार्यता की पड़ताल करती है।

थीसिस कलाई में पहने जाने वाले आईएमयू से रिकॉर्ड किए गए संकेतों का उपयोग करके एयरराइटिंग पहचान पर ध्यान केंद्रित करने के साथ शुरू होती है। एयरराइटिंग रिकग्निशन (एससीएलएआईआर) के लिए एक पर्यवेक्षित कंट्रास्टिव लॉस आधारित ढांचा प्रस्तुत किया गया है। ढांचा 2-चरण वर्गीकरण विधि को नियोजित करता है: पहले चरण में, पर्यवेक्षित विपरीत हानि को लागू किया जाता है, इसके बाद दूसरे चरण में क्रॉस-एन्ट्रॉपी हानि को कम किया जाता है। यह दृष्टिकोण सार्वजनिक रूप से उपलब्ध डेटासेट और प्रयोगशाला में दर्ज डेटासेट दोनों पर वर्गीकरण सटीकता को बढ़ाता है। इसके अतिरिक्त, एयरराइटिंग पहचान को प्रदर्शित करने वाला एक वास्तविक समय प्रदर्शन भी प्रदान किया जाता है। इसके बाद, समय-श्रृंखला डेटा (ईएमएयर) की छवि प्रस्तुति का उपयोग एयरराइटिंग पहचान के लिए भी किया जाता है। सबसे पहले, समय-श्रृंखला डेटा को छवि प्रतिनिधित्व के लिए एन्कोड किया जाता है और फिर लिखित वर्णमाला की पहचान के लिए गहन शिक्षण-आधारित मॉडल का उपयोग किया जाता है। छवि एन्कोडिंग के लिए सेल्फ सिमिलरिटी मैट्रिक्स, ग्रामियन एंगुलर फील्ड और मार्कोव ट्रांजिशन फील्ड जैसी विभिन्न तकनीकों का उपयोग किया जाता है। फिर छवि वर्गीकरण के लिए विभिन्न मानक मॉडल आर्किटेक्चर का उपयोग वर्गीकरण के लिए किया जाता है।

अग्रबाहु की मांसपेशियों से रिकॉर्ड किए गए ईएमजी संकेतों का उपयोग करके वायुलेखन पहचान की संभावना का बाद में पता लगाया गया है। अंग्रेजी के बड़े अक्षरों को लिखते समय ईएमजी संकेतों से युक्त सर्फमायोएयर डेटासेट का निर्माण किया गया है। एयरराइटिंग पहचान के लिए गहन शिक्षण मॉडल में इनपुट बनाने के लिए ईएमजी लिफाफे और अलग-अलग समय-आवृत्ति छवि प्रतिनिधित्व के निर्माण के लिए अलग-अलग समय-डोमेन सुविधाओं का पता लगाया जाता है। इस कार्य के लिए कई अलग-अलग गहन शिक्षण आर्किटेक्चर का उपयोग किया जाता है। इसके अलावा, ईएमजी-आधारित एयरराइटिंग रिकग्निशन (ट्रिप-सीईएआईआर) के लिए एक बहु-नुकसान न्यूनतमकरण ढांचा प्रस्तावित है। फ्रेमवर्क का उद्देश्य एक फीचर एम्बेडिंग वेक्टर सीखना है जो ट्रिपलेट हानि को कम करता है, साथ ही संबंधित वर्णमाला को पहचानने के लिए क्लासिफायरियर हेड के पैरामीटर सीखना भी है। प्रस्तावित विधि सर्फमायोएयर डेटासेट पर मान्य है। ट्रिपलेट लॉस, ट्रिपलेट माइनिंग रणनीतियों और फीचर एम्बेडिंग आयाम के विभिन्न रूपों के प्रभाव की भी व्यापक जांच की गई है।

एयरराइटिंग पहचान के लिए ईईजी संकेतों का लाभ उठाना मानव-कंप्यूटर इंटरैक्शन के लिए एक आशा-जनक वैकल्पिक इनपुट विधि प्रदान करता है और इस थीसिस के एक भाग के रूप में इसका पता लगाया गया है। अंग्रेजी के बड़े अक्षरों को लिखते समय रिकॉर्ड किए गए ईईजी संकेतों से युक्त न्यूरोएआईआर डेटासेट का निर्माण सबसे पहले किया गया है। फिर सटीक वायुलेखन पहचान प्राप्त करने के लिए विभिन्न गहन शिक्षण मॉडलों के संयोजन में विभिन्न विशेषताओं का पता लगाया जाता है। इन सुविधाओं में संसाधित ईईजी डेटा, स्वतंत्र घटक विश्लेषण (आईसीए) घटक, स्रोत-डोमेन-आधारित स्काउट समय श्रृंखला, और गोलाकार और हेड हार्मोनिक अपघटन-आधारित विशेषताएं शामिल हैं। इसके अलावा, सिस्टम प्रदर्शन पर विभिन्न ईईजी आवृत्ति बैंड के प्रभाव की व्यापक जांच की जाती है। यह भविष्य की प्रगति के लिए एक मजबूत आधार रेखा निर्धारित करता है और ईईजी-आधारित एयरराइटिंग मान्यता की व्यवहार्यता और उपयोगिता को प्रदर्शित करता है।

Contents

Certificate	ii
Acknowledgement	iii
Abstract	v
List of Figures	xii
List of Tables	xvi
List of Abbreviations	xx
1 Introduction	1
1.1 Literature Review and Motivation	3
1.1.1 Airwriting recognition using wrist-worn IMU	3
1.1.2 EMG-based airwriting recognition	4
1.1.3 EEG-based airwriting recognition	4
1.2 Problem Statements and Research Objectives	5
1.3 Summary of Contributions	6
1.4 Organization of the Thesis	9
2 SCLAiR: Supervised Contrastive Learning for User and Device Independent Airwriting Recognition	12
2.1 Related Work	13
2.2 Objectives and Contributions	13

2.3	Proposed Method	14
2.3.1	Problem Description	14
2.3.2	Supervised Contrastive Loss	16
2.3.3	Model Architecture	16
2.4	Experiments and Results	17
2.4.1	Datasets	17
2.4.2	Experimental Details	19
2.4.3	Results and Comparison	19
2.5	Real time demonstration	26
2.6	Summary and Contributions	27
3	ImAiR: Airwriting Recognition framework using Image Representation of IMU Signals	28
3.1	Related Work	29
3.2	Objectives and Contributions	30
3.3	Proposed Method	30
3.3.1	Problem Description	31
3.3.2	Image Encoding Methodologies	32
3.3.2.1	Self Similarity Matrix	32
3.3.2.2	Gramian Angular Field	32
3.3.2.3	Markov Transition Field	33
3.3.3	Model Architecture	34
3.4	Experiments and Results	34
3.4.1	Datasets	34
3.4.2	Experimental Details	35
3.4.3	Results and Comparison	35
3.5	Summary and Contributions	39
4	SurfMyoAiR: A surface Electromyography based framework for Airwriting Recognition	40
4.1	Related Work	41

4.2	Objectives and Contributions	43
4.3	EMG Data Collection	44
4.4	EMG Preprocessing	46
4.5	Time domain Analysis Approaches	50
4.5.1	Mean Absolute Envelope	50
4.5.2	Energy Envelope	50
4.5.3	Variance Envelope	50
4.5.4	Root Mean Square Envelope	51
4.5.5	Absolute Temporal Moment Envelope	51
4.5.6	log Detector Envelope	52
4.6	Time-Frequency Analysis Approaches	52
4.6.1	Short-Time Fourier Transform	52
4.6.2	Continuous Wavelet Transform	53
4.7	Deep Learning Frameworks	55
4.8	Experimental Setup and Results	57
4.8.1	Experimental Details	57
4.8.2	Metrics	58
4.8.3	Results	59
4.8.4	Discussion	64
4.9	Summary and Contributions	68
5	TripCEAiR: A Multi-Loss minimization approach for surface EMG based Airwriting Recognition	70
5.1	Related Work	71
5.2	Objectives and Contributions	72
5.3	Methodology	72
5.3.1	Proposed Framework	73
5.3.1.1	N-pair based triplet loss	74
5.3.1.2	Margin based triplet loss with L^2 norm	74
5.3.1.3	Margin based triplet loss with cosine similarity	75

5.3.2	Triplet Mining Strategies	76
5.3.2.1	Easy positive mining	76
5.3.2.2	Hard positive mining	76
5.3.2.3	Semihard positive mining	77
5.3.2.4	Negative Mining	77
5.3.3	Model Architecture	78
5.4	Experiments and Results	80
5.4.1	Dataset Description	80
5.4.2	Experimental Details	81
5.4.3	Results and Discussion	87
5.5	Summary and Contributions	88
6	NeuroAiR: Deep Learning Framework for Airwriting Recognition from Scalp-recorded Neural Signals	90
6.1	Related Work	91
6.2	Objectives and Contributions	93
6.3	Materials and Methods	94
6.3.1	EEG Data Recording	95
6.3.2	EEG Processing	97
6.3.2.1	Preliminary preprocessing	97
6.3.2.2	ICA decomposition	99
6.3.2.3	Artifact removal	99
6.3.2.4	EEG Filtering	100
6.3.3	Source-domain Features	100
6.3.3.1	Forward Problem	102
6.3.3.2	Inverse Problem	102
6.3.4	Spherical Harmonic Decomposition based Features	104
6.3.5	Deep Learning Frameworks	106
6.3.5.1	EEGNet	106
6.3.5.2	DeepConvNet	108

6.3.5.3	ShallowConvNet	109
6.4	Experimental Details	110
6.5	Results	112
6.6	Discussion	113
6.6.1	Deep Learning Model Analysis	113
6.6.2	EEG feature analysis	114
6.6.3	Harmonic Decomposition Analysis	116
6.6.4	Number of ICA Components analysis	117
6.6.5	EEG frequency band analysis	117
6.6.6	Comparison with existing methods	118
6.6.7	Combination of ICA components and EEGNet	119
6.7	Summary and Contributions	119
7	Conclusions and Future Directions	121
7.1	Conclusions	121
7.2	Future Directions	123
	References	123
	Appendices	140
	Appendices	141
A	FuseAiR: Fusing IMU-EMG modalities for Airwriting Recognition	142
A.1	Dataset	143
A.2	Methodology	143
A.3	Results	144
	Publications Related to Thesis Work	145
	Technical Biography of Author	146

List of Figures

1.1	Applications of wearable devices in the current scenario.	2
1.2	Block diagram depicting various modalities involved in multi-modal airwriting recognition	6
2.1	Block diagram depicting the proposed method. In the first stage an encoder with projection head based on deep network (called $Enc(\cdot)$ and $Proj(\cdot)$ respectively) is trained using supervised contrastive loss. Next, the projection head is discarded, and a linear classifier ($C(\cdot)$) is cascaded on top of the feature learner and trained with Cross Entropy loss. The dashed lines indicate that the encoders have shared weights.	15
2.2	Depiction of the data collection setup. The IMU sensor placed on the wrist is depicted in the bounding box. The user interface operated by the subject's left hand is seen on the laptop screen.	18
2.3	Mean accuracy on the target dataset using different transfer learning approaches in the unsupervised setting.	22
2.4	Mean accuracy on the target dataset using different transfer learning approaches in the supervised setting.	22
2.5	Variation of mean recognition accuracy with respect to variation in the temperature parameter (τ) and the projection encoding dimension (D_P).	24
2.6	Variation of mean recognition accuracy by varying the number of units in LSTM/BiLSTM cells.	24

2.7	Variation of accuracy by changing the number of filters (X-axis represents N_1 and the variation of N_2 is shown with different colors) and filter length for the 1DCNN architecture.	25
2.8	Scatter plot depicting clusters corresponding to each of the 26 alphabets obtained by applying dimensionality reduction using UMAP on the latent embedding vector	25
2.9	Depiction of a real-time demonstration of the IMU-based airwriting recognition system.	27
3.1	Block diagram depicting the ImAiR framework for airwriting recognition. . .	31
3.2	Recognition accuracies for LOSO experiment on Dataset-1	38
3.3	Recognition accuracies for LOSO experiment on Dataset-2	38
4.1	Depiction of the data collection setup. The sEMG electrodes and sensors placed on forearm muscles of the subject and the Tkinter based user interface are depicted.	45
4.2	Depiction of the sEMG electrode placement locations. The electrode orientation is in accordance with the muscle fibre orientation.	46
4.3	The effect of different subjects on sEMG data amplitude.	47
4.4	The effect of different repetitions on sEMG data amplitude.	47
4.5	Distribution of writing times of all samples across all subjects and alphabets. The red dashed line, black solid line and green dashed-dot lines depict the mean, median and 99.9 percentile of the lengths respectively.	48
4.6	Example of a five-channel EMG signal for the alphabet "A", interpolated to a length of 3 seconds (6000 samples) using linear interpolation (I(a)-I(e)), and the different envelopes extracted from the signal : Mean Absolute Envelope (II(a)-II(e)), Energy Envelope (III(a)-III(e)), Variance Envelope (IV(a)-IV(e)), Root Mean Squared Energy Envelope (V(a)-V(e)), Temporal Moment 3 Envelope (VI(a)-VI(e)), Temporal Moment 4 Envelope (VII(a)-VII(e)), Temporal Moment 5 Envelope (VIII(a)-VIII(e)), and log Detector Envelope (IX(a)-IX(e)).	49

4.7	Example of a five-channel EMG signal for the alphabet "A", interpolated to a length of 3 seconds (6000 samples) using linear interpolation (I(a)-I(e)), and the corresponding Time-Frequency images obtained by using Short-Time Fourier Transform (II(a)-II(e)) and Continuous Wavelet Transform (III(a)-III(e)). . .	53
4.8	Effect of variation of different parameters on mean recognition accuracies for the airwriting recognition system using MA Envelope and 1DCNN model. The left plot depicts the variation of accuracy with window length while keeping the signal length fix to 8000 samples (4 seconds), while the plot on the right depicts the variation of accuracy with signal length while keeping the window length fix to 250 samples (125 milliseconds). Effect of different interpolation techniques is marked by the legend in the subplot.	62
4.9	Effect of variation of different parameters on mean recognition accuracies for the airwriting recognition system using STFT and 2DCNN model. The left plot depicts the variation of accuracy with window length while keeping the signal length fix to 8000 samples (4 seconds), while the plot on the right depicts the variation of accuracy with signal length while keeping the window length fix to 250 samples (125 milliseconds). Effect of different interpolation techniques is marked by the legend in the subplot.	62
4.10	Confusion matrices corresponding to the MA Envelope and 1DCNN model in user-dependent scenario	63
4.11	Confusion matrices corresponding to the MA Envelope and 1DCNN model in user-independent scenario	63
5.1	Block diagram depicting the proposed method. Encoder and classifier blocks are represented by $E(\cdot)$ and $C(\cdot)$, respectively. The model parameters are optimized by minimizing L_{trip} and L_{CE} simultaneously in an end-to-end manner. The dashed lines indicate that the encoders have shared weights. The anchor sample is denoted by x_a . Corresponding positive and negative samples are denoted by x_p and x_n , respectively.	73

5.2	Depiction of the sEMG signal recording setup, the electrode placement locations, and the visual stimulus presented to the participant.	80
5.3	Effect of temperature parameter on recognition accuracy for sEMG based airwriting recognition in user dependent and independent scenarios.	84
5.4	Scatter plot depicting clusters for the different alphabets in the embedding space.	84
5.5	Confusion matrices corresponding to Semihard positive and Hard negative mining with $L_{trip-NP} + L_{CE}$ loss in the user-independent setting.	86
5.6	Confusion matrices corresponding to Semihard positive and Hard negative mining with $L_{trip-NP} + L_{CE}$ loss in the user-dependent setting.	87
6.1	Block diagram depicting an overview of the proposed EEG-based airwriting recognition system.	94
6.2	Depiction of the EEG-based airwriting data collection setup.	96
6.3	Time sheet depicting the EEG source localization plots for equally spaced time intervals between 0 to 2 seconds.	101
6.4	Depiction of the coordinate axis convention used for Spherical and Head Harmonic Decomposition in the current study.	103
6.5	The geometric depiction of a human head of radius 10 cm and perimeter 40 cm.	103
6.6	Effect of variation of the number of ICA Components on the mean recognition accuracies for the EEG-based airwriting recognition system using the EEGNet model. The accuracies for individual subjects and the mean accuracy are depicted.	115
A.1	Depiction of the networks for fusing IMU-EMG modalities.	143

List of Tables

2.1	Mean recognition accuracy for LOSO experiment on source dataset	20
2.2	Mean recognition accuracy for LOSO experiment on target dataset	20
2.3	Mean recognition accuracy on target dataset in the unsupervised setting . . .	20
2.4	Mean recognition accuracy on target dataset in the supervised setting	20
2.5	Mean accuracy on target dataset 2 without fine-tuning	21
2.6	Mean accuracy on target dataset 2 with fine-tuning	21
2.7	The 5 most confusing for LOSO experiments on source dataset.	23
2.8	The 5 most confusing for LOSO experiments on target dataset.	23
2.9	The 5 most confusing target pairs with the model trained on the source data	23
2.10	The 5 most confusing target pairs with model fine-tuned on target data . . .	23
3.1	Performance details of the used IMU in Dataset-2.	34
3.2	Recognition accuracies using different Image Encoding techniques and model architecture combinations on Dataset-1. Entries in blue denote an accuracy greater than 85% and bold entries denote the best achieved accuracy for each image encoding technique.	36
3.3	Comparison of mean recognition accuracies of the proposed ImAiR framework with state-of-the-art results for LOSO experiments on Dataset-1 and Dataset-2. Here, CE refers to Cross Entropy Loss and SCL refers to Supervised Contrastive Loss	37
4.1	A brief review of airwriting recognition systems.	41
4.2	A brief review of sEMG-based gesture recognition systems.	43

4.3	Details of the 1DCNN based classifier.	54
4.4	Details of the LSTM based classifier.	54
4.5	Details of the Bidirectional LSTM based classifier.	54
4.6	Details of the 1DCNN-LSTM based classifier.	55
4.7	Details of the 1DCNN-BiLSTM based classifier.	55
4.8	Details of the 2DCNN based classifier.	56
4.9	Recognition accuracies for the airwriting recognition task by using different time-domain sEMG envelopes and model architecture combinations in user-dependent and user-independent settings. Values are indicated as (mean $\pm$ standard deviation) of accuracies across the 5 folds.	60
4.10	Recognition accuracies for the airwriting recognition task by using different time-frequency images and model architecture combinations in user-dependent and user-independent settings. Values are indicated as (mean $\pm$ standard deviation) of accuracies across the 5 folds.	61
4.11	Alphabet-wise precision, recall and F1-score for 1DCNN model with Mean Absolute envelope.	65
4.12	The 5 most confusing alphabet pairs for MA Envelope and 1DCNN model in the user-independent setting.	66
4.13	The 5 most confusing alphabet pairs for MA Envelope and 1DCNN model in the user-dependent setting.	66
4.14	p-value of the comparative one tailed t-test between different time domain envelopes and the 1DCNN based classifier in user-independent setting. Entries with a significance value below 0.05 are depicted in blue.	67
4.15	p-value of the comparative one tailed t-test between different time domain envelopes and the 1DCNN based classifier in user-dependent setting. Entries with a significance value below 0.05 are depicted in blue.	68
5.1	Details of the proposed 1DCNN-BiLSTM based encoder.	78
5.2	Details of the 1DCNN based encoder.	79
5.3	Details of the Stacked LSTM-based encoder.	79

5.4	Details of the Stacked Bidirectional LSTM-based encoder.	79
5.5	Details of the 1DCNN-LSTM based encoder.	80
5.6	Recognition accuracies for the user-independent airwriting recognition task by using different triplet mining strategies and feature embedding dimension. Entries in blue denote best achieved accuracy for each embedding dimension and the best overall accuracy is depicted in bold.	82
5.7	Recognition accuracies for the user-dependent airwriting recognition task by using different triplet mining strategies and feature embedding dimension. Entries in blue denote best achieved accuracy for each embedding dimension and the best overall accuracy is depicted in bold.	83
5.8	Alphabet-wise precision, recall and F1-score corresponding to Semihard positive and Hard negative mining with $L_{trip-NP} + L_{CE}$ loss in user-independent, and user-dependent evaluation settings.	85
5.9	Comparison of different model architectures for sEMG based airwriting recognition task.	86
6.1	Details of the demographic details of the subjects involved in data collection.	96
6.2	Details of the EEGNet model architecture.	107
6.3	Details of the DeepConvNet model architecture.	108
6.4	Details of the ShallowConvNet model architecture.	109
6.5	Mean recognition accuracies for the EEG-based airwriting recognition task by using Processed EEG, ICA Components, and Scout time series. The effect on accuracies by using different frequency bands in conjunction with different Deep Learning models is also depicted.	111
6.6	Mean recognition accuracies for the EEG-based airwriting recognition task by using Spherical Harmonic Decomposition based features $[V_{nm}^{SH}]$. The effect on accuracies by using different decomposition order and frequency bands in conjunction with different Deep Learning models is also depicted.	111

6.7	Mean recognition accuracies for the EEG-based airwriting recognition task by using Head Harmonic Decomposition based features $[V_{nm}^{H^2}]$. The effect on accuracies by using different decomposition order and frequency bands in conjunction with different Deep Learning models is also depicted.	112
6.8	p-values of the comparative one-tailed t-test between different deep learning models. Entries with a significance value below 0.05 are depicted in bold. . .	114
6.9	p-values of the comparative one-tailed t-test between different features used for EEG-based airwriting recognition using the EEGNet model. Entries with a significance value below 0.05 are depicted in bold.	114
6.10	p-values of the comparative one-tailed t-test between different frequency bands used for EEG-based airwriting recognition using the EEGNet model and ICA source activity matrix. Entries with a significance value below 0.05 are depicted in bold.	115
6.11	Prediction time analysis for different features.	117
6.12	Performance of the proposed models for EEG-Based airwriting recognition in comparison to state-of-the-art methods.	118
A.1	Classification accuracies using the proposed method.	144

List of Abbreviations

ALP	Associated Legendre Polynomial
BCI	Brain-Computer Interface
BEM	Boundary Element Method
BiLSTM	Bidirectional Long Short Term Memory
CE	Cross Entropy
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DANN	Domain-Adversarial Training of Neural Networks
DeepJDOT	Deep Joint Distribution Optimal Transport
DRCN	Deep Reconstruction-Classification Networks
DTW	Dynamic Time Warping
ECoG	Electrocorticography
EEG	Electroencephalography
EMD	Emperical Mode Decoposition
EMG	Electromyography
FFT	Fast Fourier Transform
GADF	Gramian Angular Difference Field
GAF	Gramian Angular Field
GASF	Gramian Angular Summation Field
HCI	Human-Computer Interaction
HHD	Head Harmonic Decomposition
ICA	Independent Component Analysis
IMU	Inertial Measurement Unit
LSTM	Long Short Term Memory
MAV	Mean Absolute Value
MRI	Magnetic Resonance Imaging
MTF	Markov Transition Field

ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SCL	Supervised Contrastive Loss
sEMG	surface Electromyography
SHD	Spherical Harmonic Decomposition
sLORETA	standard Low-Resolution Electrical Tomography
SSM	Self Similarity Matrix
STFT	Short-Time Fourier Transform
tanh	hyperbolic tangent
UMAP	Uniform Manifold Approximation and Projection