EXPLAINABLE METRIC LEARNING FOR SMALL-SIZED MEDICAL DATASETS

Shipra Madan

Department of Electrical Engineering
Indian Institute of Technology-Delhi

August 2024

EXPLAINABLE METRIC LEARNING FOR SMALL-SIZED MEDICAL DATASETS

by

Shipra Madan

Department of Electrical Engineering

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

Indian Institute of Technology-Delhi

August 2024

Dedicated to

My Children : Kabir & Veer

Certificate

This is to certify that the thesis entitled **"Explainable Metric Learning for small-sized Medical Datasets"** being submitted by **Ms. Shipra Madan** 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 the bonafide research work carried out by her 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.

(Prof. Santanu Chaudhury)
Dept. of Electrical Engineering
Indian Institute of Technology-Delhi
Hauz Khas, New Delhi 110016
INDIA

(Prof. Tapan Kumar Gandhi)
Dept. of Electrical Engineering
Indian Institute of Technology-Delhi
Hauz Khas, New Delhi 110016
INDIA

Acknowledgements

First and foremost, I extend my deepest gratitude to the good GOD Almighty, whose boundless grace and guidance have made this thesis possible. In every step of this journey, I have felt their presence, and I am humbled by the wisdom and strength they have bestowed upon me.

As I stand on the cusp of completing this transformative journey, I am profoundly grateful to the individuals and institutions that have contributed to the realization of this thesis. This thesis marks the culmination of countless hours of research, collaboration, and growth, and I am humbled by the support that has brought it to fruition.

First and foremost, I extend my heartfelt appreciation to my advisors, Prof. Santanu Chaudhury and Prof. Tapan Gandhi, for their unwavering guidance, invaluable insights, and continuous support throughout the entirety of my research. Their mentorship has shaped not only the direction of this thesis but also my growth as a scholar.

I am indebted to the members of my thesis committee, Prof. Sumantra Dutta Roy, Prof. Brejesh Lall, and Prof. Anup Singh, for their constructive feedback, thoughtful discussions, and expertise that have enriched the quality of this work.

The vibrant academic environment at the Indian Institute of Technology Delhi (IITD) has been the crucible in which this research has flourished. The resources, mentorship, and collaborative ethos of the Electrical Department have been indispensable in my growth as a researcher.

I am thankful for the financial support provided by Ministry of Human Resource Development, India (MHRD), which allowed me to dedicate my efforts to this research and to attend conferences that expanded my horizons.

My sincere appreciation goes to my peers and colleagues who have kindled stimulating debates, provided feedback, and shared their insights, enriching the academic discourse.

The unwavering support of my family has been my bedrock. Their belief in my potential and their enduring encouragement have been my wellspring of inspiration.

Finally, I would like to express my heartfelt gratitude to my friends who have provided moments of respite, laughter, and camaraderie, reminding me of the importance of balance and shared experiences.

This thesis is the embodiment of the collaborative efforts, intellectual camaraderie, and unyielding support of numerous individuals and entities. The above acknowledgments only begin to scratch the surface of the profound impact you all have had on this work.

Thank you for being part of this remarkable journey.

Shipra Madan

Shipra Madan

Abstract

In recent years, Deep Learning has garnered significant attention within the research community, owing to its remarkable performance across a spectrum of tasks. Consequently, deep learning models have found successful applications in diverse domains, including but not limited to speech recognition, autonomous vehicle navigation, medical image analysis, disease diagnosis, as well as broader tasks like segmentation and classification. Despite the widespread adoption of deep learning, there remains ample room for improvement, particularly in scenarios where data availability is limited. Additionally, deep learning models are often regarded as enigmatic "black-box" systems, making it challenging to elucidate the rationale behind the labels they assign. This thesis centers on addressing two critical areas: first, the effective handling of small datasets through the utilization of the deep metric learning paradigm, and second, the development of methods for explaining the decisions made by neural networks trained with limited data. These endeavors are specifically tailored to the domain of medical imaging.

To begin, I present the notion of metric learning, an approach centered on the use of a distance metric to discern the similarity or dissimilarity between images. This technique leverages a limited number of images from each category to facilitate classification. In this context, I employ Siamese networks and Triplet nets, both stemming from the metric learning paradigm, to construct classification models tailored for small-sized medical datasets. Some of the medical problems which are tackled in various experimental studies in this work are bone-age assessment in 0-3 years of age, COVID-19 detection, pneumonia classification and lung cancer prediction.

Subsequently, the thesis shifts its focus towards crafting interpretable frameworks, de-

veloped through the metric learning approach, and trained on compact medical datasets. These frameworks not only yield meaningful insights but also come accompanied by comprehensive explanations that elucidate the underlying mathematical models. To commence this exploration, I provide a concise introduction to Explainable Artificial Intelligence (XAI) techniques, delving into their extensive categories applicable in the realm of deep learning for the analysis of medical images. Following that, I outline the methodology employed to explain the models that have been trained on the aforementioned medical datasets with limited data, employing various explanation methods, including post-hoc, model-specific, model-agnostic, local, and counterfactual techniques.

In the later part of the thesis, a systematic comparison of Explainable Artificial Intelligence (XAI) methods with their suitability of application to different dataset scenarios is discussed. To evaluate both the quantitative and qualitative effectiveness of each interpretation method introduced in this study, we have devised two comprehensive validation metrics.

In summary, these frameworks pave the way to not only create new classification frameworks which work for small-scale medical datasets but also provide logical explanations to the decisions made by models trained using small datasets.

सार

नवीनतम वर्षों में, विभिन्न कार्यों में अपने उल्लेखनीय प्रदर्शन के कारण, डीप लर्निंग ने अनुसंधान समुदाय के भीतर महत्वपूर्ण ध्यान आकर्षित किया है। नतीजतन, डीप लर्निंग मॉडल को विभिन्न डोमेन में सफल अनुप्रयोग मिले हैं, जिनमें भाषण पहचान, स्वायत्त वाहन नेविगेशन, चिकित्सा छवि विश्लेषण, रोग निदान, साथ ही विभाजन और वर्गीकरण जैसे व्यापक कार्य शामिल हैं। डीप लर्निंग को व्यापक रूप से अपनाने के बावजूद, सुधार की पर्याप्त गुंजाइश बनी हुई है, खासकर उन परिदृश्यों में जहां डेटा उपलब्धता सीमित है। इसके अतिरिक्त, डीप लर्निंग मॉडल को अक्सर रहस्यमय "ब्लैक-बॉक्स" सिस्टम के रूप में माना जाता है, जिससे उनके द्वारा दिए गए लेबलों के पीछे का कारण समझना कठिन हो जाता है। यह थीसिस दो महत्वपूर्ण क्षेत्रों पर ध्यान केंद्रित करती है: पहला, डीप मीट्रिक लर्निंग प्रतिमान के उपयोग के माध्यम से छोटे डेटासेट का प्रभावी संचालन, और दूसरा, सीमित डेटा के साथ प्रशिक्षित न्यूरल नेटवर्क द्वारा किए गए निर्णयों को समझाने के तरीकों का विकास। ये प्रयास विशेष रूप से मेडिकल इमेजिंग के क्षेत्र के अनुरूप हैं।

प्सबसे पहले , मैं मीट्रिक सीखने की धारणा प्रस्तुत करती हूँ , जो छवियों के बीच समानता या असमानता को समझने के लिए दूरी माप का उपयोग करती है। इस तकनीक ने प्रत्येक श्रेणी से एक सीमित संख्या की छवियों का उपयोग कर क्लासिफिकेशन को सुविधाजनक बनाने में मदद की है। इस संदर्भ में, मैं मेट्रिक लर्निंग परिदृश्य से उत्पन्न दो मॉडल्स, सियामी नेटवर्क और ट्रिपलेट नेट्स का उपयोग करती हूँ । इस कार्य में विभिन्न प्रायोगिक अध्ययनों में जिन कुछ चिकित्सीय समस्याओं का समाधान किया गया है उनमें 0-3 वर्ष की आयु में हड्डी आयु का मूल्यांकन, कोविड-19 (COVID-19) पता लगाना, न्यूमोनिया वर्गीकरण और फेफड़ों के कैंसर की पूर्वानुमान शामिल हैं।

इसके बाद, थीसिस अपना ध्यान व्याख्यात्मक रूपरेखा तैयार करने की ओर केंद्रित करती है, जिसे मीट्रिक शिक्षण दृष्टिकोण के माध्यम से विकसित किया जाता है, और कॉम्पैक्ट मेडिकल डेटासेट पर प्रशिक्षित किया जाता है। ये ढाँचे न केवल सार्थक अंतर्दृष्टि प्रदान करते हैं बल्कि व्यापक स्पष्टीकरण के साथ आते हैं जो अंतर्निहित गणितीय मॉडल को स्पष्ट करते हैं। इस अन्वेषण को शुरू करने के लिए, मैं चिकित्सा छवियों के विश्लेषण के लिए डीप लर्निंग के क्षेत्र में लागू होने वाली उनकी व्यापक श्रेणियों पर प्रकाश डालते हुए व्याख्या योग्य कृत्रिम बुद्धिमत्ता (एक्सएआई) तकनीकों का एक संक्षिप्त परिचय प्रदान करती हूँ। इसके बाद, मैं उन मॉडलों को समझाने के लिए नियोजित कार्यप्रणाली की रूपरेखा तैयार करती हूँ, जिन्हें सीमित डेटा के साथ उपरोक्त मेडिकल डेटासेट पर प्रशिक्षित किया गया है, जिसमें पोस्ट-हॉक, मॉडल-विशिष्ट, मॉडल-अज्ञेयवादी, स्थानीय और प्रतितथ्यात्मक तकनीकों सहित विभिन्न स्पष्टीकरण विधियों को नियोजित किया गया है।

थीसस के अंतम भाग में , विभिन्न डेटासेट परिदृश्यों में उनके अनुप्रयोग की उपयुक्तता के साथ व्याख्या योग्य कृत्रिम बुद्धिमत्ता (एक्सएआई) विधियों की एक व्यवस्थित तुलना पर चर्चा की गई है। इस अध्ययन में प्रस्तुत प्रत्येक व्याख्या पद्धति की मात्रात्मक और गुणात्मक प्रभावशीलता दोनों का मूल्यांकन करने के लिए, हमने दो व्यापक सत्यापन मेट्रिक्स तैयार किए हैं।

संक्षेप में, ये ढाँचे न केवल नए वर्गीकरण ढाँचे बनाने का मार्ग प्रशस्त करते हैं जो छोटे पैमाने के चिकित्सा डेटासेट के लिए काम करते हैं बल्कि छोटे डेटासेट का उपयोग करके प्रशिक्षित मॉडलों द्वारा लिए गए निर्णयों को तार्किक स्पष्टीकरण भी प्रदान करते हैं।

Contents

List of Figures

List of Tables

Abbreviations

AP	Antero Posterior Supine
AUC	Area Under Curve
AUROC	Area Under the receiver operator Curve
BAA	Bone Age Assessment
CAD	Computer Aided Design
CAM	Class Activation Mapping
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Networks
COVID19	Corona Virus Disease
CT	Computed Tomography
CXR	Chest X-ray
DeepLIFT	Deep Learning Important FeaTures
DHA	Digital Hand Atlas Database
DICOM	Digital Imaging and Communications in Medicine
DML	Distance Metric Learning
DNN	Deep Neural Networks
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GAN	Generative Adverserial Networks
GP	Greulich–Pyle
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
Grad-CAM	Gradient-weighted Class Activation Mapping

GRU	Gated Recurrent Unit
IITD	Indian Institute of Technology-Delhi
IQ-OTH/NCCD	Iraq-Oncology Teaching Hospital/National Center for Cancer Disease
LIME	Local Interpretable Model Agnostic Explanation
LRP	Layerwise Relevance Propagation
MERS-Co	Middle East Respiratory Syndrome Coronavirus
MLP	Multi Layer Perceptron
MRI	Magnetic Resonance Imaging
PCA	Principal Component Analysis
RISE	Randomized Input Sampling for Explanation
RNN	Recurrent Neural Networks
ROC	Reciever Operating Characterstic
ROI	Regions of Interest
RSNA	Radiological Society of North America
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SARS	Severe Acute Respiratory Syndrome
SGD	Stochastic Gradient Descent
TN	True Negative
TP	True Positive
TPR	True Positive Rate
tSNE	t-Distributed Stochastic Neighbour Embedding
TW	Tanner–Whitehouse
WHO	World Health Organization
XAI	Explainable Artifical Intelligence