Indian Institute of Technology Delhi

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ADVANCING ENERGY SYSTEM MODELING AND CYBERSECURITY THROUGH HOLISTIC AI DEVELOPMENT

ARNAB BHATTACHARJEE

DOCTOR OF PHILOSOPHY

INDIAN INSTITUTE OF TECHNOLOGY DELHI

&

THE UNIVERSITY OF QUEENSLAND

JULY 2024

Indian Institute of Technology Delhi

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ADVANCING ENERGY SYSTEM MODELING AND CYBERSECURITY THROUGH HOLISTIC AI DEVELOPMENT

by

ARNAB BHATTACHARJEE

Submitted

in fulfillment of the requirements for the joint degree of

DOCTOR OF PHILOSOPHY

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

&

THE UNIVERSITY OF QUEENSLAND

JULY 2024

Dedicated to Chance, as it is all that is.

Supervisor Certification

This is to certify that the thesis entitled "**Advancing Energy System Modelling and Cybersecurity through Holistic AI development**" being submitted by **Mr. Arnab Bhattacharjee** to the Indian Institute of Technology Delhi and The University of Queensland for the award of degree of Doctor of Philosophy is a record of *bonafide* research work carried out by them. **Mr. Arnab Bhattacharjee** has worked under our guidance and supervision and has fulfilled the requirements for the submission of this thesis, which to our knowledge has reached the requisite standard. The results contained in this thesis are original and have not been submitted, in part or full, to any other University or Institute for the award of any other degree or diploma.

Prof. Ashu Verma
Professor, DESE
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi
India 110016

Prof. Sukumar Mishra
Professor, EE
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi
India 110016

Prof. Tapan Kumar Saha
Professor, EECS
The University of Queensland
St. Lucia, Brisbane
Australia, QLD 4072

Abstract

The contemporary energy landscape is undergoing unprecedented shifts, propelled by the integration of renewable energy resources and the emergence of smart grid technologies. This evolution challenges the traditional power system paradigms rooted in synchronous machine-dominated grids. While the dependence of renewable energy generation on ambient conditions introduces uncertainties and non-dispatchability on the generation side, the increased penetration of inverter-based generators, loads, and energy storage devices is leading to unprecedented changes in the power network and demand sides in terms of network strength, demand characteristics, directionality of energy flow and the level of uncertainties associated with them. Two major problems rooting from these changes are addressed in this work – a. the increased analytical intractability in accurate energy system components and network-level modelling, and b. growing susceptibility of smarter grid technologies to evasive cyberattacks.

Through three broad objectives, this thesis evaluates the limitations of conventional power system techniques in accommodating increased complexities, specifically in energy system modelling and cybersecurity tasks. It further proposes the development of holistic Artificial Intelligence and Machine Learning (AI/ML) as alternative models to address these limitations. Accordingly, the chapters in this thesis cater to individual objectives and can be perceived as standalone works in terms of the specific energy system application they are addressing. What connects them is the underlying theme of identifying common limitations in the conventional power system strategies and the subsequent development of holistic AI models as better alternatives.

While AI/ML development for power system applications has received a lot of interest in recent years, the existence of challenges like difficulty in generalization to black-box operating conditions, unavailability of human-understandable interpretations and recourse for model behaviour, large-scale annotated data requirement, susceptibility to adversarial attacks and privacy threats make the deployability of AI/ML models in cyber-physical spaces questionable. This thesis addresses these challenges by first laying down a skeletal notion of 'holistic' AI by defining favourable properties expected in AI models employed in cyber-physical systems. Explicit strategies to induce relevant holistic properties in the developed AI models for the different objectives are discussed and their implementation through architectural modifications and newer training paradigms are highlighted.

In the first objective of the thesis, the role of AI in modelling analytically intractable processes in modern energy systems is explored. Specifically, it deals with developing reliable Li-ion battery models focusing on three key aspects of battery modelling - current state estimation, lifetime response modelling, and lifetime degradation modelling. The intractability of these processes and the limitations of existing low-dimensional modelling paradigms are first elaborated. Further, a multi-scale temporal convolutional neural network is proposed for current battery state estimation and a very long-term

sequence modeling approach is proposed for battery lifetime modeling. Relevant comparative analysis is carried out with the existing state of the art (SOTA) which validates the superior performance of the developed models in terms of modelling accuracy and generalization characteristics by catering to a multitude of different black box operating conditions and battery specifications. The properties of generalizability and interpretability for recourse are incorporated in the proposed models through novel architecture development and training paradigms.

The subsequent parts of the thesis investigate the cyber-threat landscape of smart energy systems and highlight the challenges faced by existing energy system models and advanced AI/ML models employed for attack detection and mitigation. Firstly, the rapidly advancing threat landscape of cyber-physical systems leading to increased possibilities of evasive injection attacks characterized by high impact and low visibility is explored. While advanced AI/ML detectors exist to isolate such attacks, they require large amounts of annotated data representing attacked and non-attacked system conditions and lack generalizability to different attack models. In the second objective, this thesis addresses some of these challenges by proposing an unsupervised deep latent space clustering algorithm for isolating highly stealthy injection attacks against the nonlinear state estimation operation of a power system. The proposed model shows significant improvements compared to the existing state of the art in terms of stealthy attack detection performance, generalizes equally well to different attack models, doesn't require annotated data, and demonstrates robustness against a class of grey-box adversarial attacks.

Finally, a pre-emptive threat model exploration against cyber-physical systems is carried out in the third objective to identify minimal system information requiring cyber-attacks that can target the disruption of key functions, not necessarily relying on AI models, while retaining their adversarial nature against numerous black box AI and non-AI systems that are closely related to the target operation. Accordingly, a Deep Black Box Adversarial Attack, acronym DeeBBAA, is developed that targets the standard non-linear state estimation operation of an unknown power system while consistently evading detection from a multitude of state-of-the-art black box anomaly detectors with widely different operating principles.

The proposed thesis advocates for paradigm shifts in energy system modelling, management, and cybersecurity, endorsing the integration of Artificial Intelligence (AI) and its derivatives as potential solutions. AI's adaptability to handle diverse, unstructured data and its model-free function approximation capabilities hold promise in addressing these intricate challenges. However, while AI presents promising solutions, this thesis acknowledges its limitations. Challenges such as the need for extensive labelled datasets, lack of interpretability of models, cybersecurity concerns, and ensuring broad applicability across varying operating conditions are highlighted and possible solutions are proposed. Further, the threats posed and vulnerabilities associated with advancements in AI on critical infrastructure are explored to motivate the development of necessary pre-emptive measures and stronger defences.

The long-term goal of this research is to continue exploring and developing AI-based methodologies that transcend the constraints of traditional power system analysis, offering robust, adaptable, and cyber-resilient energy management solutions. It calls for collaborative efforts to advance AI techniques, navigate data accessibility and security concerns, and ensure ethical deployment in transforming the energy landscape towards a sustainable, resilient future.

सार

समकालीन ऊर्जा परिदृश्य अभूतपूर्व बदलावों से गुजर रहा है, जो नवीकरणीय ऊर्जा संसाधनों के एकीकरण और स्मार्ट ग्रिड प्रौद्योगिकियों के उद्भव से प्रेरित है। यह विकास पारंपरिक पावर सिस्टम प्रतिमानों को चुनौती देता है जो सिंक्रोनस मशीन-प्रभुत्व वाले ग्रिडों पर आधारित हैं। जबकि नवीकरणीय ऊर्जा उत्पादन की पर्यावरणीय स्थितियों पर निर्भरता पीढ़ी के पक्ष में अनिश्चितता और गैर-प्रेषणीयता को पेश करती है, इन्वर्टर-आधारित जनरेटर, लोड और ऊर्जा भंडारण उपकरणों की बढ़ती पैठ नेटवर्क की ताकत, मांग की विशेषताओं, ऊर्जा प्रवाह की दिशा और उनसे जुड़े अनिश्चितताओं के स्तर के संदर्भ में बिजली नेटवर्क और मांग पक्षों में अभूतपूर्व परिवर्तनों को जन्म दे रही है। इन परिवर्तनों से उत्पन्न दो प्रमुख समस्याओं का इस कार्य में समाधान किया गया है - a. सटीक ऊर्जा प्रणाली घटकों और नेटवर्क-स्तरीय मॉडलिंग में बढ़ी हुई विश्लेषणात्मक दुर्गमता, और b. स्मार्ट ग्रिड प्रौद्योगिकियों की बढ़ती संवेदनशीलता को धोखेबाज साइबर हमलों के लिए।

तीन व्यापक उद्देश्यों के माध्यम से, यह थीसिस बढ़ी हुई जटिलताओं को समायोजित करने में पारंपरिक पावर सिस्टम तकनीकों की सीमाओं का मूल्यांकन करती है, विशेष रूप से ऊर्जा प्रणाली मॉडलिंग और साइबर सुरक्षा कार्यों में। यह आगे इन सीमाओं को दूर करने के लिए वैकल्पिक मॉडल के रूप में समग्र कृत्रिम बुद्धिमत्ता और मशीन लर्निंग (AI/ML) के विकास का प्रस्ताव करती है। तदनुसार, इस थीसिस के अध्याय व्यक्तिगत उद्देश्यों को पूरा करते हैं और जिस विशिष्ट ऊर्जा प्रणाली अनुप्रयोग को वे संबोधित कर रहे हैं, उसके संदर्भ में स्वतंत्र कार्यों के रूप में देखा जा सकता है। जो उन्हें जोड़ता है वह पारंपरिक पावर सिस्टम रणनीतियों में सामान्य सीमाओं की पहचान करने और बेहतर विकल्प के रूप में समग्र एआई मॉडल के विकास की अंतर्निहित थीम है।

हालांकि पावर सिस्टम अनुप्रयोगों के लिए एआई/एमएल विकास में हाल के वर्षों में बहुत रुचि प्राप्त हुई है, ब्लैक-बॉक्स ऑपरेटिंग स्थितियों में सामान्यीकरण में कठिनाई, मानव-सुविधाजनक व्याख्याओं और मॉडल व्यवहार के लिए समाधान की अनुपलब्धता, बड़े पैमाने पर एनोटेटेड डेटा आवश्यकता, विरोधी हमलों के प्रति संवेदनशीलता और गोपनीयता खतरों जैसी चुनौतियों के अस्तित्व से साइबर-फिजिकल स्पेस में AI/ML मॉडल की तैनाती संदिग्ध हो जाती है। यह थीसिस सबसे पहले साइबर-फिजिकल सिस्टम में नियोजित एआई मॉडल में अपेक्षित अनुकूल गुणों को परिभाषित करके 'समग्र' एआई की एक कंकाल धारणा रखकर इन चुनौतियों का समाधान करती है। विभिन्न उद्देश्यों के लिए विकसित एआई मॉडलों में प्रासंगिक समग्र गुणों को प्रेरित करने के लिए स्पष्ट रणनीतियों पर चर्चा की जाती है और उनके कार्यान्वयन को आर्किटेक्चरल संशोधनों और नए प्रशिक्षण प्रतिमानों के माध्यम से उजागर किया जाता है।

थीसिस के पहले उद्देश्य में, आधुनिक ऊर्जा प्रणालियों में विश्लेषणात्मक रूप से दुर्गम प्रक्रियाओं के मॉडलिंग में एआई की भूमिका का पता लगाया गया है। विशेष रूप से, यह बैटरी मॉडलिंग के तीन प्रमुख पहलुओं - वर्तमान स्थिति अनुमान, जीवनकाल प्रतिक्रिया मॉडलिंग, और जीवनकाल अपक्षय मॉडलिंग पर ध्यान केंद्रित करते हुए विश्वसनीय लिथियम-आयन बैटरी मॉडल विकसित करने से संबंधित है। इन प्रक्रियाओं की दुर्गमता और मौजूदा कम-आयामी मॉडलिंग प्रतिमानों की सीमाओं को पहले

विस्तृत किया गया है। आगे, वर्तमान बैटरी स्थिति अनुमान के लिए एक बहु-स्केल अस्थायी कन्वोल्यूशनल न्यूरल नेटवर्क प्रस्तावित किया गया है और बैटरी जीवनकाल मॉडलिंग के लिए एक बहुत लंबी-अवधि अनुक्रम मॉडलिंग दृष्टिकोण प्रस्तावित किया गया है। विभिन्न ब्लैक बॉक्स ऑपरेटिंग स्थितियों और बैटरी विशिष्टताओं के लिए सेवा प्रदान करके मॉडलिंग सटीकता और सामान्यीकरण विशेषताओं के संदर्भ में विकसित मॉडलों के श्रेष्ठ प्रदर्शन को मान्य करने वाले मौजूदा सर्वोत्तम स्थिति (SOTA) के साथ प्रासंगिक तुलनात्मक विश्लेषण किया जाता है। प्रस्तावित मॉडलों में सामान्यता और पुनर्विचार के लिए व्याख्यात्मकता के गुणों को उपन्यास आर्किटेक्चर विकास और प्रशिक्षण प्रतिमानों के माध्यम से शामिल किया गया है।

थीसिस के बाद के भागों में स्मार्ट ऊर्जा प्रणालियों के साइबर-खतरे परिदृश्य की जांच की गई है और हमले का पता लगाने और शमन के लिए नियोजित मौजूदा ऊर्जा प्रणाली मॉडलों और उन्नत एआई/एमएल मॉडलों द्वारा सामना की जाने वाली चुनौतियों को उजागर किया गया है। सबसे पहले, साइबर-फिजिकल सिस्टम के तेजी से उन्नत खतरे परिदृश्य की खोज की गई है जिससे उच्च प्रभाव और कम दृश्यता वाली अत्यधिक चोरी की इंजेक्शन हमलों की संभावनाएं बढ़ जाती हैं। जबकि ऐसे हमलों को अलग करने के लिए उन्नत एआई/एमएल डिटेक्टर मौजूद हैं, उन्हें हमले और गैर-हमले की प्रणाली की स्थितियों का प्रतिनिधित्व करने वाले बड़े मात्रा में एनोटेटेड डेटा की आवश्यकता होती है और विभिन्न हमले मॉडलों के लिए सामान्यीकरण की कमी होती है। दूसरे उद्देश्य में, यह थीसिस गैर-रैखिक राज्य अनुमान ऑपरेशन के खिलाफ अत्यधिक चोरी के इंजेक्शन हमलों को अलग करने के लिए एक असुपरवाइज्ड डीप लेटेंट स्पेस क्लस्टरिंग एल्गोरिदम का प्रस्ताव करके इन चुनौतियों में से कुछ का समाधान करती है। प्रस्तावित मॉडल, चोरी हमले का पता लगाने के प्रदर्शन के संदर्भ में मौजूदा सर्वोत्तम स्थिति के मुकाबले महत्वपूर्ण सुधार दिखाता है, विभिन्न हमले मॉडलों के लिए समान रूप से सामान्यीकृत करता है, एनोटेटेड डेटा की आवश्यकता नहीं होती है, और ग्रे-बॉक्स विरोधी हमलों के एक वर्ग के खिलाफ मजबूती का प्रदर्शन करता है।

अंत में, तीसरे उद्देश्य में साइबर-फिजिकल सिस्टम के खिलाफ एक पूर्व-खतरे मॉडल अन्वेषण किया गया है ताकि न्यूनतम प्रणाली जानकारी की पहचान की जा सके जो साइबर-हमले की आवश्यकता हो सकती है जो प्रमुख कार्यों के विघटन को लक्षित कर सकते हैं, जरूरी नहीं कि एआई मॉडलों पर निर्भर हो, जबकि व्यापक रूप से संबंधित ब्लैक बॉक्स एआई और गैर-एआई सिस्टम के खिलाफ उनके विरोधी स्वभाव को बनाए रखते हुए। तदनुसार, एक डीप ब्लैक बॉक्स एडवर्सेरियल अटैक, संक्षेप में डीब्बा, विकसित किया गया है जो अज्ञात पावर सिस्टम के मानक गैर-रैखिक राज्य अनुमान ऑपरेशन को लक्षित करता है जबकि व्यापक रूप से अलग-अलग ऑपरेटिंग सिद्धांतों के साथ सर्वोत्तम स्थिति ब्लैक बॉक्स विसंगति डिटेक्टरों की एक बहुलता से लगातार बचाव करता है।

प्रस्तावित थीसिस ऊर्जा प्रणाली मॉडलिंग, प्रबंधन और साइबर सुरक्षा में प्रतिमान बदलावों की वकालत करती है, जो कृत्रिम बुद्धिमत्ता (एआई) और इसके व्युत्पन्न को संभावित समाधान के रूप में एकीकृत करती है। विविध, असंरचित डेटा को संभालने के लिए एआई की अनुकूलता और इसके मॉडल-फ्री फ़ंक्शन अनुमान क्षमताएं इन जटिल चुनौतियों को संबोधित करने में संभावनाएं रखती हैं। हालांकि, जबकि एआई आशाजनक समाधान प्रस्तुत करता है, यह थीसिस इसकी सीमाओं को स्वीकार करती है। चुनौतियां जैसे

व्यापक लेबल वाले डेटा की आवश्यकता, मॉडलों की व्याख्यात्मकता की कमी, साइबर सुरक्षा चिंताएं, और विभिन्न ऑपरेटिंग स्थितियों में व्यापक लागू योग्यता सुनिश्चित करना उजागर किया गया है और संभावित समाधानों का प्रस्ताव किया गया है। आगे, एआई में प्रगति के कारण महत्वपूर्ण बुनियादी ढांचे के साथ जुड़े खतरों और कमजोरियों का पता लगाने के लिए आवश्यक पूर्व-खतरे उपायों और मजबूत रक्षा प्रणाली के विकास को प्रेरित करने के लिए अन्वेषण किया गया है।

इस शोध का दीर्घकालिक लक्ष्य पारंपरिक पावर सिस्टम विश्लेषण की बाधाओं को पार करने वाले एआई-आधारित पद्धतियों का अन्वेषण और विकास करना जारी रखना है, मजबूत, अनुकूलनशील, और साइबर-लचीले ऊर्जा प्रबंधन समाधानों की पेशकश करना। यह एआई तकनीकों को आगे बढ़ाने, डेटा पहुंच और सुरक्षा चिंताओं को नेविगेट करने, और एक स्थायी, लचीले भविष्य की ओर ऊर्जा परिदृश्य को बदलने में नैतिक तैनाती सुनिश्चित करने के लिए सहयोगात्मक प्रयासों का आह्वान करता है।

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications included in this thesis

[1] **Arnab Bhattacharjee**, Arnab Kumar Mondal, Ashu Verma, Sukumar Mishra and Tapan K. Saha, Deep Latent Space Clustering for Detection of Stealthy False Data Injection Attacks Against AC State Estimation in Power Systems, *IEEE Transactions on Smart Grid*, vol. 14, no. 3, May 2023.
[2] **Arnab Bhattacharjee**, Sukumar Mishra and Ashu Verma, Deep Adversary based Stealthy False Data Injection Attacks against AC state estimation, *2022 IEEE PES 14th Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2022.
[3] **Arnab Bhattacharjee**, Ashu Verma, Sukumar Mishra and Tapan K. Saha, Estimating State of Charge for xEV Batteries Using 1D Convolutional Neural Networks and Transfer Learning, *IEEE Transactions on Vehicular Technology* vol. 70, no. 4, April 2021

Submitted manuscripts included in this thesis

[4] **Arnab Bhattacharjee**, Guangdong Bai, Wayes Tushar, Tapan K. Saha, Ashu Verma, and Sukumar Mishra, DeeBBAA: A benchmark Deep Black Box Adversarial Attack against Cyber-Physical Power Systems, *arXiv, Under Minor Revision in IEEE Journal of Internet of Things*, 2023.

Other publications during candidature

[5] Arnab Kumar Mondal, **Arnab Bhattacharjee**, Parag Singla and A. P. Prathosh, xViTCOS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography, *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, 2022.
[6] Arnab Mondal, **Arnab Bhattacharjee**, Sudipto Mukherjee, Himanshu Asnani, Sreeram Kannan, Prathosh A P, C-MI-GAN : Estimation of Conditional Mutual Information using MinMax formulation, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, PMLR 124:849-858, 2020.

Contributions by others to the thesis

All Chapters

Editorial suggestions were provided by the candidate's supervisors, Prof. Tapan K. Saha (UQ), Prof. Sukumar Mishra (IIT Delhi), and Prof. Ashu Verma (IIT Delhi) and PhD committee comprising Prof. Bijaya Ketan Panigrahi (IIT Delhi), Prof. Nilanjan Senroy (IIT Delhi) and Dr. Rahul Sharma (UQ).

Statement of parts of the thesis submitted to qualify for the award of another degree

No works submitted towards another degree have been included in this thesis.

Research involving human or animal subjects

No animal or human subjects were involved in this research.

Acknowledgments

This thesis is a compilation of the work I have carried out during my Ph.D. at the Indian Institute of Technology Delhi, India (IITD) and The University of Queensland, Australia (UQ). I would like to express my deepest gratitude to my supervisors, Prof. Tapan Kumar Saha, Prof. Sukumar Mishra, and Prof. Ashu Verma for making me a part of their research teams and for their unconditional support throughout my Ph.D. Their experience, knowledge, and research aptitudes have been key factors for fuelling my motivation for research. Their patience, constant efforts, and dedication have played a key role in shaping my professional and personal life.

I would also like to thank my ASMC members, Prof. B.K. Panigrahi, Prof. Nilanjan Senroy, and Dr. Rahul Sharma for their valuable feedback on my work. Their constructive comments and invaluable advice have resulted in the enhancement of the presented work. I express my heartfelt gratitude towards Dr. Wayes Tushar for being a constant source of inspiration whenever I needed him and for the impromptu brainstorming sessions, we had during my time at UQ. My heartfelt gratitude goes to the UQ-IITD Academy of Research team for greatly facilitating the journey of my PhD in two premier institutions of the world.

My deepest regards to Dr. A.P. Prathosh and Mr. Arnab Kumar Mondal for holding my hands into the field of academic writing and publications. I wouldn't have been where I am, had I not met you in the very first year of my Ph.D.

Then there are people who have been with me these five arduous years during my darkest nights and my brightest days. I extend my sincerest love and appreciation to Arjita Pal, Dr. Mohammad Abdullah Bareen, and B.V. Suryakiran for lending me your ears and hearts whenever I needed them the most. Thanks just for existing. Also, we couldn't have survived Australia had it not been for one of the most pro-active and organised souls that very fortunately came along with us, the one and only Dr. Imon Chakraborty. Australia was a wonderful experience, more so because I crafted some refreshing bonds of friendships with some very young and talented people - Zhixuan Li, your hotpots are thoroughly missed.

PhDs get stressful very often and some people and relations are lost in that midst. To those who tried their best to bear with me and lost, thanks for trying your best.

I also extend my gratitude and best wishes to the friends, colleagues, and seniors who I got acquainted with in my classes and workspaces – Bevin K.C., Sonam Mittal, Anuj, Ankit Singh, Manish Kumar Yadav, Gaurav Prit, Nikilesh Wanjari, Mohd. Juned, Arinjoy Biswas, Dr. Sumedha Sharma, Dr. Himanshu Grover, Dr. Neha Tak and Dr. Ram Krishan. Thank you for your support and being a part of my PhD journey.

Finally, there is a saying that goes, "you may leave home, but home never leaves you". To my home – my parents, Mrs. Nilima Bhattacharjee and Mr. Prabhat Ranjan Bhattacharjee, who went against all odds to raise the person that I am today and my annoying but irreplaceable elder brother, Mr Arabinda Bhattacharjee and my wonderful sister-in-law Mrs. Srimanti Paul, I extend boundless gratitude and love. Thanks for being my rocks and having my back.

Life is a manifestation of combined effort. For those who couldn't be named but played even the smallest parts in making my life better in ways I can't even fathom, thank you for being there.

Financial support

This research was supported jointly by an Indian Institute of Technology Delhi Research Scholarship, the Prime Minister's Research Fellowship (PMRF), and Research grants from the UQ-IITD Academy of Research (UQIDAR) and PMRF.

Keywords

Artificial Intelligence, Deep Learning, Holistic AI, Unsupervised Learning, Li-ion Battery, State Estimation, Cyber-Physical Security, Adversarial Attacks, False Data Injection Attacks

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 090608, Renewable Power and Energy Systems Engineering, 40%
ANZSRC code: 080109, Pattern Recognition and Data Mining, 40%
ANZSRC code: 080110, Simulation and Modelling, 20%

Fields of Research (FoR) Classification

FoR code:0801, Artificial Intelligence and Image Processing, 60%
FoR code: 0906, Electrical and Electronic Engineering, 40%

Contents

Supervisor Certification **i**

Abstract ii

Abstract in Hindi vi

Contents **xv**

List of Figures **xix**

List of Tables **xxiv**

List of Abbreviations and Symbols **xxv**

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Thesis Proposition 3

1.2.1 Thesis Objectives 5

1.2.2 Overview of proposed objectives 6

1.3 Holistic AI development for energy applications 8

1.3.1 Extrapolative Generalization 10

1.3.2 Interpretability 12

1.3.3 Robustness to Adversarial Threats 13

1.3.4 Privacy 14

1.3.5 Holistic AI development goals 15

1.4 Overview of Thesis 16

2 Literature Review **17**

2.1 Li-ion battery Modelling 18

2.1.1 Motivation 18

2.1.2 Battery State Estimation 21

2.1.3 Existing Model Classes 22

2.1.4 Summary and Relevance to Objective A 36

2.2 Data Integrity Attacks in Energy Systems 36

2.2.1 Motivation . . . 36
2.2.2 Background . . . 38
2.2.3 SFDI attacks - Categories and Design Aspects . . . 40
2.2.4 Detection of SFDIA . . . 41
2.2.5 Summary and Relevance to Objective B . . . 45
2.3 Vulnerability Assessment of AI . . . 46
2.3.1 Susceptibility to Adversarial Attacks . . . 46
2.3.2 Adversarial attacks against cyber-physical systems . . . 49
2.3.3 Summary and Relevance to Objective C . . . 51

3 Generalizable Li-ion battery State of Charge Estimation 53
3.1 Introduction . . . 53
3.2 Proposed Method . . . 55
3.2.1 Problem Statement . . . 55
3.2.2 CNN for SoC Estimation . . . 56
3.2.3 Regularization in CNN model . . . 57
3.2.4 Knowledge Transfer across datasets . . . 59
3.3 Dataset description . . . 60
3.3.1 Dataset A – Panasonic 18650 PF battery dataset . . . 60
3.3.2 Dataset B – LG 18650 HG2 battery dataset . . . 62
3.3.3 Additive Noise in Data . . . 62
3.4 Experimental Setup . . . 62
3.4.1 Architecture Selection . . . 62
3.4.2 Model training and evaluation . . . 63
3.4.3 Transfer learning on different dataset . . . 66
3.5 Results and Discussions . . . 68
3.5.1 Architecture Selection . . . 68
3.5.2 Model Evaluation . . . 70
3.5.3 Transfer learning on different dataset . . . 73
3.6 Summary . . . 75

4 Long-term Li-ion battery response and degradation modelling 79
4.1 Introduction . . . 79
4.2 Methodology . . . 84
4.2.1 Problem Statement . . . 84
4.2.2 CRA-aft . . . 85
4.3 Training and Inference . . . 87
4.3.1 Battery Aging Datasets . . . 88
4.3.2 Input-output formulation for model training . . . 89
4.3.3 Multi-task Learning . . . 90

4.3.4 Training Paradigms . . . 90
4.3.5 Autoregressive Inference: . . . 92
4.4 Results . . . 94
4.4.1 Comparison with baselines . . . 94
4.4.2 Li-ion Battery Response Prediction . . . 95
4.4.3 Li-ion battery Capacity Trajectory Prediction . . . 97
4.5 Relevant Discussions and Future Roadmap . . . 99
4.6 Summary . . . 101

5 Detection of Stealthy False Data Injection against Power System 103
5.1 Introduction . . . 103
5.2 Proposed Detection Strategy . . . 104
5.2.1 Stacked AutoEncoder . . . 105
5.2.2 Self Training for Improving Cluster Assignment . . . 107
5.3 Case Study . . . 108
5.3.1 Test Systems . . . 108
5.3.2 Primary Attack Model . . . 109
5.3.3 Stealthy FDIA Data Generation . . . 110
5.3.4 Deep Latent Space Clustering based Detection . . . 111
5.3.5 Performance Metrics . . . 112
5.3.6 Baselines . . . 114
5.3.7 Time Complexity . . . 117
5.3.8 Resiliency against Adversarial Attacks . . . 117
5.4 Results . . . 119
5.4.1 Detection Performance - Accuracy, Generalizability and Data efficiency . . . 119
5.4.2 Resiliency against Grey box attacks . . . 121
5.5 Summary . . . 121

6 Vulnerability Analysis of AI in Power System Cybersecurity 125
6.1 Introduction . . . 126
6.2 Adversarial Attacks - Background . . . 127
6.3 Proposed Attack Strategy . . . 128
6.3.1 Identification of Attack Region and Reconnaissance . . . 129
6.3.2 Partial Approximation of the Unknown AC-PSSE . . . 131
6.3.3 Adversarial Optimization Against Proxy State Estimator . . . 132
6.4 Simulation and Results . . . 136
6.4.1 Test System and Data Generation . . . 136
6.4.2 Training the Neural State Estimator . . . 137
6.4.3 Training the Defense Baselines . . . 138
6.4.4 Designing attack vectors using DeeBBAA . . . 142

6.4.5 Results and Inferences 144
6.5 Summary 149

7 Conclusion **157**
7.1 Summary of Individual Chapters 157
7.2 Future Scopes 161

Bibliography **163**

List of Figures

2.1 OCV curves over SoC for different Li-ion cell chemistries [7]. 19

2.2 Path dependence of LiB battery degradation [8]. 20

2.3 Microscale models are full-order mechanistic models that incorporate active non-homogenous cathode and anode particles and non-homogeneity in ion and electrolyte distributions. Homogenous models assume homogeneity over particle shapes and electrolytes. The DFN model further simplifies the 3D electrode-electrolyte geometry into a single dimension and assumes unilateral flow of ions through the electrolyte over one dimension. Non-uniformity in particle behaviour is still incorporated in DFNs. Further model reduction leads to SPM and SPMe models, which remove the particle non-uniformly by considering single representative anode and cathode particles [9]. 24

2.5 *Capacity Fade Curve of a representative LiB showing three regions of slow, intermediate, and high rates of degradation.* 28

2.6 Flow Diagram of the nonlinear AC State Estimation in Power Systems (AC-PSSE). . . . 37

2.7 Flow diagram of Conventional Bad Data Detection in Power systems. 39

2.8 Example of an adversarial perturbation on an image classification network. Adding the small perturbation causes no visual change to the image of the panda. However, the deep learning model misclassifies it as a gibbon with high probability. The perturbation is developed using Fast Gradient Sign Method [10]. 47

2.9 Pictorial representation of the steps involved in developing transfer-based black box attacks. 49

3.1 Merge first architecture 58

3.2 Dense first architecture 58

3.3 CNN Model Training Flowchart 59

3.4 CNN Model training with Transfer Learning 59

3.5 Cycle 2 drive cycle data at 25°C from the Panasonic Dataset 61

3.6 Mixed 2 drive cycle data at 25°C from the LG 18650 HG2 Dataset 61

3.7 Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the Panasonic battery dataset, sampled at 10 Hz, depending on the type of additive noise in data and history size t_w 63

3.8 Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the Panasonic battery dataset, sampled at 1 Hz, depending on the type of additive noise in data and history size t_w . . . 65
3.9 Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the Panasonic battery dataset, sampled at 0.1 Hz, depending on the type of additive noise in data and history size t_w . . . 65
3.10 Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the LG battery dataset, sampled at 10 Hz, depending on the type of additive noise in data and history size t_w . . . 66
3.11 Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the LG battery dataset, sampled at 1 Hz, depending on the type of additive noise in data and history size t_w . . . 66
3.12 Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the LG battery dataset, sampled at 0.1 Hz, depending on the type of additive noise in data and history size t_w . . . 67
3.13 Transfer Learning Results: Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the LG battery dataset after training the CNN model using transfer learning with the entire LG training data, sampled at 10 Hz. The variations are shown with respect to the type of additive noise in data and history size t_w . . . 68
3.14 Transfer Learning Results: Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the LG battery dataset after training the CNN model using transfer learning with the entire LG training data, sampled at 1 Hz. The variations are shown with respect to the type of additive noise in data and history size t_w . . . 69
3.15 Transfer Learning Results: Variations in MAE and MAX values obtained on US06 and HWFET drive cycles of the LG battery dataset after training the CNN model using transfer learning with only 40% of the training cycles. The variations are shown with respect to the type of additive noise in data and data sampling frequency. The t_w is fixed at 500. . . 69
3.16 SoC estimation on US06 drive cycle at 25°C from the Panasonic Dataset, sampled at 1 Hz, with different forms of noise. The CNN model consists of a 2C-DF architecture, and t_w was equal to 500. . . . 71
3.17 SoC estimation on US06 drive cycle at 25°C from the LG Dataset, sampled at 1 Hz, with different forms of noise. The CNN model consists of a 2C-DF architecture, and t_w was equal to 500. . . . 72
3.18 Training and Validation loss vs Epochs on Panasonic dataset, sampled at 10 Hz and with t_w = 500 and no added noise. . . . 72
3.19 Transfer Learning Results: SoC estimation on US06 drive cycle at 25°C from the LG Dataset, sampled at 1 Hz, with different forms of noise. The CNN model, trained using transfer learning with the entire LG training dataset, consists of a 2C-DF architecture and t_w was equal to 500. . . . 73

3.20 Transfer Learning Results: SoC estimation on US06 drive cycle at 25°C from the LG Dataset, sampled at 1 Hz, with different forms of noise. The CNN model, trained using transfer learning with 40% of LG training dataset, consists of a 2C-DF architecture and t_w was equal to 500. 74

3.21 Training loss convergence characteristics for a noiseless LG dataset sampled at 10 Hz and with t_w =500. Three plots correspond to cases when the CNN model is trained with and without transfer learning. The with transfer learning case is further of two types, one where the entire LG training data is used for training and the other where 40% data is used for training. 74

4.1 Three full equivalent cycles are depicted in this figure. The first highlighted in green is also a full cycle according to the original definition of a cycle, while the remaining two are formed using multiple partial cycles. 81

4.2 Detailed layout of the Chunk Autoformer Block. 86

4.3 An overview of the proposed recurrent encoder - decoder architecture CRA-aft. 87

4.4 Detailed layout of the autoregressive inference (also forced learning) mechanism. 91

4.5 Predicted vs true battery states - voltage and SoC for the first and final 5 prediction cycles using CRA-aft by autoregressive inference for an LFP battery. The first 5 prediction cycles correspond to cycles 50 -55, as the initial 50 cycles are required as inputs to fine-tune the CRA-aft encoder. 97

4.6 Predicted vs true battery states - voltage and SoC for the first and final 5 prediction cycles using CRA-aft by autoregressive inference for an NMC battery. The first 5 prediction cycles correspond to cycles 50 -55, as the initial 50 cycles are required as inputs to fine-tune the CRA-aft encoder. 97

4.7 Predicted vs true battery states - voltage and SoC for the first and final 5 prediction cycles using CRA-aft by autoregressive inference for an NCA battery. The first 5 prediction cycles correspond to cycles 50 -55, as the initial 50 cycles are required as inputs to fine-tune the CRA-aft encoder. 98

4.8 Predicted vs true battery states - voltage and SoC for the first and final 5 prediction cycles using CRA-aft by autoregressive inference for a highly irregularly cycled battery from the NASA-Rand dataset. The first five prediction cycles correspond to cycles 50 -55 as the initial 50 cycles are required as inputs to fine-tune the CRA-aft encoder. 98

4.9 Predicted vs true capacity degradation trajectory using CRA-aft by autoregressive inference for three different battery types. 99

4.10 Layout of a conceptual framework for optimal design of future battery inputs. 100

5.1 Block Diagram of the proposed method 105

5.2 Detailed operational breakdown of the proposed method. 106

5.3 Hourly load profile over a week. 109

5.4 Distribution of attack scale incurred by the optimization-based attack model in the form of the number of attacked buses and maximum state deviations caused. 112
5.5 Characteristics of the attack vectors generated using the optimization-based attack model. 113
5.6 Visualization of t-SNE projections of the ground truth data (1st column of each row), clusters obtained using k-Means baseline algorithm(2nd column of each row) and clusters obtained using the proposed method (3rd column of each row) when the imbalance ratio is approximately equal to 1:2. The three rows correspond to the IEEE 14 bus, 118 bus and 300 bus test systems respectively. The data is inseparable in the original feature space but forms clean clusters in the latent variable space. 114
5.7 Visualization of t-SNE projections of the ground truth data (1st column of each row), clusters obtained using k-Means baseline algorithm(2nd column of each row) and clusters obtained using the proposed method (3rd column of each row) when the imbalance ratio is approximately equal to 1:8. The three rows correspond to the IEEE 14 bus, 118 bus and 300 bus test systems respectively. The data is inseparable in the original feature space but forms clean clusters in the latent variable space. 115
5.8 Performance of DLSC regarding detection of stealthy grey-box FGSM attacks of varying strengths on the IEEE 118 bus system. 121

6.1 Graph representations of the IEEE 14 bus system demonstrating different types of attack regions. The red nodes represent buses in the attack region, the dotted edges indicate power lines included in the attack region, the magenta nodes and the solid edges represent buses and power lines not included in the attack region. The bus with the blue asterisk beside it represents the initial target bus for the localized case. 130
6.2 Training the Neural State Estimator f_ϕ using historical data samples. 132
6.3 Stealthy Black Box Adversarial Attack Formulation against Power System State Estimation using the proposed DeeBBAA framework. 136
6.4 Summary of data generation and subsequent purpose of usage of different datasets. Dataset A is only used to train the NSE in the offline stage of DeeBBAA. Dataset B is used for training the defence baselines and the online generation of attack vectors using DeeBBAA. This is done to ensure the worst-case scenario from the attacker's perspective, where the defence baselines are trained on the same data which is used to generate the online DeeBBAA attacks. At the same time, the NSE is trained on a completely different dataset to prevent data spillage. 138
6.5 Distribution of Largest Normalized residues corresponding to benign data samples 140
6.6 Distribution of L2 norm of residues corresponding to benign data samples 141
6.7 Distribution of eigenvalues of $\boldsymbol{W}$. 144
6.8 Probability of DeeBBAA attacks bypassing conventional LNRT and χ^2 tests for IEEE 39 bus and 118 bus systems under different attack regions, epsilons and number of compromised measurements. 144

6.9 Probability of DeeBBAA attacks bypassing statistical consistency and learning-based defences for IEEE 39 bus system under localized attack region with different values of epsilons and the number of compromised measurements. 146
6.10 Probability of DeeBBAA attacks bypassing statistical consistency and learning-based defences for IEEE 39 bus system under delocalized attack region with different values of epsilons and the number of compromised measurements. 146
6.11 Probability of DeeBBAA attacks bypassing statistical consistency and learning-based defences for IEEE 118 bus system under localized attack region with different values of epsilons and the number of compromised measurements. 147
6.12 Probability of DeeBBAA attacks bypassing statistical consistency and learning-based defences for IEEE 118 bus system under delocalized attack region with different values of epsilons and the number of compromised measurements. 147
6.13 Deviation statistics for IEEE 39 bus test system. The number of measurements compromised by the attack vector is plotted on the x-axis, and the y-axis consists of the deviations. For each of the values in the x-axis, there are 4 boxplots, each corresponding to a value of ε as shown in the legends. Subplots a, b, c, and d correspond to the localized attack region setting and plots e, f, g, and h correspond to the delocalized attack region case. (a,e). Maximum Deviation in real power measurements, (b,f). Maximum deviation in reactive power measurements, (c,g). Median Deviation in Voltage magnitude estimates (in log scale), (d,h). Median deviation in voltage angle estimates (in log scale). 152
6.14 Deviation statistics for the IEEE 118 bus test system. The number of measurements compromised by the attack vector is plotted on the x-axis, and the y-axis consists of the deviations. For each of the values in the x-axis, there are 4 boxplots, each corresponding to a value of ε as shown in the legends. Subplots a, b, c, and d correspond to the localized attack region setting and plots e, f, g, and h correspond to the delocalized attack region case. (a,e). Maximum Deviation in real power measurements, (b,f). Maximum deviation in reactive power measurements, (c,g). Median Deviation in Voltage magnitude estimates (in log scale), (d,h). Median deviation in voltage angle estimates (in log scale). 153
6.15 Point Deviation for the IEEE 39 and 118 bus test systems. 154

List of Tables

3.1 Architecture of the best-performing model obtained from the architecture selection experiment . . . 64
3.2 Estimation error for the architecture selection experiment. The errors under each drive cycle are in the form MAE (%) / MAX(%) . . . 70
3.3 SoC estimation error values obtained on the test datasets derived from the Panasonic Battery dataset for the $2C-DF$ architecture. The errors under each drive cycle are in the form MAE (%) / MAX(%). f_s stands for the sampling frequency of data. . . . 70
3.4 SoC estimation error values obtained on the test datasets derived from LG Battery dataset for the $2C-DF$ architecture. The errors under each drive cycle are in the form MAE (%) / MAX(%). f_s stands for the sampling frequency of data. . . . 70
3.5 Transfer Learning Results: SoC estimation error values obtained on the test datasets derived from the LG Battery dataset for the $2C-DF$ architecture with transfer learning. Results for both cases, one where the entire LG training data is used and the other where 60% of the training data is stochastically eliminated, are given. The errors under each drive cycle are in the form MAE (%) / MAX(%). f_s stands for the sampling frequency of data. . . . 71

4.1 Comparison of CRA-aft on long-term time series forecasting with state-of-the-art . . . 95
4.2 Li-ion battery lifetime response prediction results using CRA-aft encoder . . . 96
4.3 Li-ion battery lifetime capacity degradation trajectory prediction results using CRA-aft encoder-decoder . . . 99

5.1 Parameters for FDIA attack generation . . . 109
5.2 Comparison with Baselines . . . 119
5.3 Comparison of SFDIA detection performance of unsupervised baselines on different attack models and under different imbalances . . . 120

6.1 Details of Attack Regions for the IEEE 39 and 118 bus test systems . . . 139
6.2 Details of defensive baselines used to test the evasive properties of the SFDIA generated using DeeBBAA. . . . 151

List of Abbreviations and Symbols

Abbreviations	
AC+TD	AutoCorrelation and Time Delay Aggregation
AEMO	Australian Energy Market Operator
AI	Artificial Intelligence
ARI	Adjusted Rand Index
ARMA	Autoregressive Moving Average
BDD	Bad Data Detection
BMS	Battery Management System
BPTT	Backpropagation Through Time
CNN	Convolutional Neural Networks
CPS	Cyber-Physical System
CRA-aft	Chunk Recurrent Autoformer with adaptive forced training
DeeBBAA	Deep Black Box Adversarial Attack
DER	Distributed Energy Resource
DFN	Doyle Fuller Newman
DL	Deep Learning
DLSC	Deep Latent Space Clustering
DoD/DOD	Depth of Discharge
DRL	Deep Reinforcement Learning
E3LM	Enhanced Extreme Learning Machine
ECM	Equivalent Circuit Model
EIS	Electrochemical Impedance Spectroscopy
EoL/EOL/EoUL	End of (Useful) Life
ERM	Empirical Risk Minimization
EV	Electric Vehicle
FDI	False Data Injection
FEC	Full Equivalent Cycle
FOM	Fractional Order Model
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model

GNN	Graph Neural Network
HaiD/HAiD	Holistic AI Development
HWFET	Highway Fuel Economy Driving Schedule
IC	Incremental Capacity
IOM	Integer Order Model
IR	Imbalance ratio
KLD	Kullback-Liebler Divergence
LFP	Lithium Ferrous Phosphate
LiB	Li-ion Battery
LNRT	Largest Normalized Residue Test
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multi Layer Perceptron
MPC	Model Predictive Control
MSE	Mean Squared Error
MTL	Multi-Task Learning
NCA	Nickel Cobalt Aluminium oxide
NMC	Nickel Manganese Cobalt oxide
NMI	Normalized Mutual Information
NSE	Neural State Estimator
NYISO	New York Independent System Operator
OCV	Open Circuit Voltage
PCA	Principal Component Analysis
PCDM	Perturbation Constrained Deviation Maximization
PMU	Phasor Measurement Units
PSSE	Power System State Estimation
QCQP	Quadratic Constrained Quadratic Programming
ReLU	Rectified Linear Unit
RL	Reinforcement Learning
RNN	Recurrent Neural Networks
RPT	Reference Performance Tests
RUL	Remaining Useful Life
S4	Structured State Space Sequence
SAE	Stacked Autoencoder
SCADA	Supervisory Control and Data Acquisition
SDP	Semi Definite Programming
SEI	Solid Electrolyte Interface
SFDI	Stealthy False Data Injection
SoC/SOC	State of Charge
SoH/SOH	State of Health

SOTA	State of the art
SPM	Single Particle Model
SPMe	Single Particle Model with electrolyte
SSL	Self Supervised Learning
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
TL	Transfer Learning
t-SNE	t-Students' distributed Neigbourhood Embeddings
UDDS	Urban Dynamometer Driving Schedule
UPF	Unscented Particle Filter
US06	Supplemental Federal Test Procedure Driving Schedule
V2G	Vehicle to Grid
V2V	Vehicle to Vehicle
etc.	*etc.*