

Preface

The rapid advancements in Machine Learning (ML) and High Performance Computing (HPC) have catalyzed a transformative era in scientific research and technology. As editors of the “Handbook of Research on Integrating Machine Learning Into HPC-Based Simulations and Analytics,” we are honored to present a comprehensive collection that bridges these two powerful domains, showcasing innovative research and pioneering applications.

The impetus for this handbook arises from a significant challenge in contemporary research: the limitations of small-scale computing systems in handling the vast and rapidly growing datasets characteristic of grand-challenge problems. Traditional ML models, while effective in many domains, are often constrained by the memory and processing resources available. This has led to an urgent need for integrating ML with HPC environments to not only enhance scalability but also to improve the performance and efficiency of predictive models.

In recent years, the remarkable strides in ML, particularly deep learning, have underscored the necessity for substantial computational power. HPC systems, with their superior processing capabilities, provide an ideal platform to meet this demand. The synergy between ML and HPC promises to accelerate progress in various research areas by enabling faster data processing, more complex simulations, and ultimately, more profound insights.

This book is a response to the evolving technological landscape and the growing interdependence of ML and HPC. It is designed to serve as a repository of cutting-edge research, practical solutions, and insightful discussions on the intersection of these fields. Our aim is to highlight the sustained and broad efforts invested by the research community, illustrating how ML and HPC can complement and enhance each other.

This collection of chapters in the book delves into cutting-edge advancements in artificial intelligence, machine learning, and high-performance computing, exploring their applications across diverse domains. From innovative insult detection on social media and explainable AI techniques to domain adaptation in scene classification and fitness activity recognition, the chapters highlight the transformative potential of these technologies. Further discussions include optimizing deep learning for mobile applications, leveraging edge computing for data lakes, and integrating AI with IoT for energy management, health-care, and nuclear research.

Chapter 1 addresses the growing issue of cyberbullying and online abusive language, presenting a Transformer-based deep learning model for insult detection on social media platforms. Leveraging the BERT method, the approach focuses on text pre-processing, representation, and supervised classification, achieving superior results compared to traditional CNN and RNN models, with a macro F1-score of 0.83 and an accuracy of 86%.

Exploring the intersection of machine learning and explainability, this Chapter 2 examines robust variable selection methods to minimize the influence of data outliers. It compares prior, intrinsic, and posterior approaches using datasets from gene expression, crime rates, and tourism infrastructure, showcasing how explainable models improve decision-making in contaminated datasets.

Chapter 3 introduces the Semi-Supervised Domain Adaptation Network (SSDAN) for remote sensing scene classification, highlighting its ability to reduce data labeling requirements. With over 99% accuracy and Kappa coefficient values in various domain scenarios, the method demonstrates its efficiency in leveraging multi-source data for improved classification performance.

Focusing on fitness activity tracking, chapter 4 explores deep learning models that map smartphone sensor data into predefined activity classes. Using benchmark datasets, the models deliver high accuracy in recognizing patterns, aiding users in monitoring performance and calorie expenditure with enhanced precision.

To address memory and energy constraints in mobile applications, chapter 5 examines DNN compression techniques, such as pruning and redesigning neural architectures. It evaluates these methods on metrics like compression rate, inference time, and accuracy, offering guidance for practitioners to optimize mobile DNN performance.

Chapter 6 evaluates the role of High-Performance Edge Computing (HPEC) in processing big data for deep learning applications. By employing data lake frameworks and platforms like Jetson Nano, it highlights advancements in efficiency and computational performance, achieving classification speeds as fast as 8.19 ms per image with 90% accuracy.

Chapter 7 advocates integrating machine learning and high-performance computing into electric machine design. It reviews traditional cascade algorithms and demonstrates the potential of ML and optimization techniques to overcome design challenges, paving the way for adaptable and efficient methodologies.

Exploring the integration of HPC and distributed deep learning frameworks, chapter 8 delves into designing scalable models using libraries like TensorFlow, Keras, and Horovod. It examines their deployment on platforms like AWS and GCP, optimizing real-time data analytics for reliable and functional neural networks.

Chapter 9 explores how ML and HPC collaboration accelerates discoveries in fields like materials science, climate modeling, and drug discovery. Case studies highlight the ability of these technologies to automate analytics, optimize resource use, and solve global scientific challenges with unprecedented accuracy.

Focusing on renewable energy and environmental monitoring, chapter 10 explores the integration of HPC and AI for optimizing energy systems and analyzing environmental data. It showcases case studies demonstrating enhanced precision and efficiency in sustainable energy harvesting and ecosystem management.

Chapter 11 surveys AI techniques like reinforcement learning for optimizing energy consumption in HPC systems. By dynamically adapting workloads, these methods achieve energy efficiency without compromising computational performance, contributing to sustainability goals in high-demand environments.

Integrating federated learning and gestural technology, chapter 12 demonstrates advancements in patient care and diagnostics. Use cases in telemedicine and rehabilitation illustrate how decentralized ML models and human-computer interaction technologies streamline workflows while maintaining data privacy.

Chapter 13 highlights the transformative role of advanced visual information processing technologies in neuroscience and healthcare. By enhancing diagnostics and treatment personalization, it demonstrates how deep learning and computer vision revolutionize neural and cognitive research while addressing ethical and data challenges.

Focusing on food safety and waste reduction, chapter 14 examines the application of microcontrollers and IoT-enabled systems for real-time food quality monitoring. Innovations in sensor technology and predictive analytics ensure precision in managing temperature, humidity, and contamination in supply chains.

Chapter 15 explores the application of IoT devices and ML algorithms in nuclear research. It discusses how real-time data collection and predictive modeling enhance precision in analyzing nuclear reactions, addressing challenges related to data security and computational demands.

By integrating HPC into electric vehicle design, chapter 16 examines battery management, energy efficiency, and thermal dynamics. It discusses real-time simulations and prototyping that optimize EV systems, reduce development costs, and advance sustainability in the automotive industry.

Chapter 17 reviews cutting-edge HPC architectures for ML workloads, emphasizing GPU, TPU, and FPGA technologies. It discusses distributed computing frameworks like Apache Spark and future trends like quantum computing, offering insights into improving scalability, energy efficiency, and computational performance.

The book's objectives are multifaceted: to present recent advancements in ML techniques applied on HPC systems, to discuss significant results achieved, and to identify ongoing challenges and future opportunities at this intersection. By fostering collaboration between the ML and HPC communities, we hope to promote rapid and seamless advancements towards an ecosystem that benefits both disciplines.

The convergence of ML and HPC is not merely a technical endeavor but a strategic move to overcome existing barriers related to application complexity and machine cost. To this end, this handbook sets four primary goals:

1. To develop optimized solutions that significantly reduce computational requirements.
2. To provide a dynamic and comprehensive view of domain sciences.
3. To foster interdisciplinary collaboration for integrated knowledge development.
4. To stimulate innovations with deep societal impact through advancements in scientific research.

Our target audience includes academics, researchers, computer scientists, and engineers, as well as experts in ML, HPC, and data analytics. Practitioners in computational science, artificial intelligence, and data science, along with graduate and undergraduate students interested in these fields, will find this book particularly valuable.

The topics covered are diverse, reflecting the broad scope of this interdisciplinary field:

- High Performance Computing systems
- Machine Learning paradigms
- Parallelization and scaling of ML techniques/algorithms
- ML applications on HPC systems
- HPC system design and optimization for ML workloads
- Convergence of HPC and deep learning
- Data analytics using ML on HPC systems

- Networking and storage solutions for ML on HPC systems
- Libraries, tools, and workflows for ML on HPC systems
- Emerging trends in HPC and ML

We are grateful to the contributors whose expertise and dedication have made this handbook possible. Their collective efforts provide a rich source of knowledge and inspiration for anyone engaged in the fascinating journey of integrating ML into HPC-based simulations and analytics.

With great anticipation, we present this handbook, confident that it will serve as an essential resource for advancing research and practice at the confluence of ML and HPC.

Belgacem Ben Youssef

King Saud University, Saudi Arabia

Mohamed Maher Ben Ismail

King Saud University, Saudi Arabia