Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
https://doi.org/10.36930/40340613
Article received 28.08.2024
Article accepted 05.09.2024
UDC 004.[056.5+02+71+77+85+89]
Correspondence author: O. P. Kuziv oleksii.p.kuziv@lpnu.ua
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)

O. P. Kuziv, M. A. Nazarkevych
Lviv Polytechnic National University, Lviv, Ukraine

OPTIMIZATION OF MACHINE LEARNING MODEL TRAINING PROCEDURE ON MULTI-GPU SYSTEMS TO ENHANCE CYBER SECURITY IN TELECOMMUNICATION NETWORKS

This paper analyzes the optimization features of machine learning (ML) model training procedures using multi-GPU systems to enhance cyber security in telecommunication networks. A key aspect of the study is the use of data parallelism, which allows the distribution of the training load across multiple GPUs, significantly reducing training time and improving model accuracy-critical factors for rapid threat detection in cyberspace. A novel approach for optimizing data batch size using Mutual Information (MI) is proposed, which harmonizes the utilization of computational resources with the information content of the training data. MI helps to determine the optimal data batch size that minimizes training errors and improves model accuracy without a significant increase in training time. Experimental results demonstrate the substantial advantages of multi-GPU configurations compared to single-GPU setups, providing faster training and improved model accuracy. It was particularly emphasized that MI-guided batch size tuning significantly outperforms traditional manual tuning methods, ensuring higher validation accuracy and reducing training time. The study showed that the MI-based approach is an effective tool for addressing the problem of optimizing ML model training processes in real-world scenarios where cyber security is critical. The proposed methods allow ML models to train faster and more accurately identify potential threats, making them particularly relevant for telecommunication networks where a rapid response to new threats in real time is required. The implementation of modern computational technologies such as multi-GPU systems and MI-optimized training enhances the efficiency and accuracy of machine learning models. This, in turn, improves cyber security measures and ensures a more reliable defence of telecommunication networks against malicious attacks. It is noted that the proposed approaches can be adapted not only for cyber security but also for other areas where high model accuracy and fast training are important. Future research prospects include the development of new machine learning methods, particularly deep neural networks, the exploration of alternative computational architectures such as quantum computing or distributed systems, and their integration into real-time systems. Special attention should be paid to the ethical aspects of implementing automated cyber security systems, particularly in preventing bias in algorithms and ensuring fairness in their application.

Keywords: parallel processing; training procedure optimization; cyber threat detection; machine learning in cyber security; mutual information.

Introduction / Вступ

In the realm of telecommunication networks, cyber security stands as a critical defense against the rising wave of digital threats that have increased both in complexity and volume. The advent of the digital age has introduced new connectivity paradigms, creating a landscape where cyber-attacks target the very core of communication infrastructures. These attacks disrupt services, endanger data integrity, and violate privacy, making the protection of telecommunication networks a paramount concern for the modern world [3, 9]. Against this backdrop, the application of machine learning (ML) has emerged as a promising solution. Specifically, ML techniques, such as image recognition, present novel ways to detect and mitigate sophisticated cyber threats, many of which cannot be effectively addressed by traditional security measures [10, 18, 19].

Relevance of Research: with the rapid evolution of cyber threats, there is an urgent need for more robust cyber security solutions. ML models offer dynamic capabilities for detecting threats and responding in real time, but their development and deployment face significant computational challenges. In particular, multi-GPU configurations offer an avenue to optimize ML training, improving both efficiency and accuracy [14, 16, 17]. This research is thus crucial in advancing the state of the art in cyber security within telecommunication networks, where the stakes of a successful defense are high.

Scientific Context: the rapid expansion of cyber threats in telecommunication networks necessitates the exploration of advanced ML techniques to ensure these networks' security. While numerous studies have explored the use of ML in various cyber security domains, multi-GPU setups represent an underexplored field with great potential to enhance the speed and precision of ML training [5, 21]. Given that

Інформація про авторів:
Кузів Олексій Павлович, магістр, аспірант, кафедра інформаційних мереж та систем.
Email: oleksii.p.kuziv@lpnu.ua; https://orcid.org/0009-0007-9365-2487
Назаркевич Марія Андріївна, д-р техн. наук, професор, кафедра інформаційних мереж та систем.
Email: mariia.a.nazarkevych@lpnu.ua; https://orcid.org/0000-0002-6528-9867

Цитування за ДСТУ: Кузів О. П., Назаркевич М. А. Оптимізація процедури машинного навчання моделі на багатопроцесорних установках для посилення кібербезпеки в телекомунікаційних мережах. Науковий вісник НЛТУ України. 2024, т. 34, № 6. С. 93–100.
Citation APA: Kuziv, O. P., & Nazarkevych, M. A. (2024). Optimization of machine learning model training procedure on multi-gpu systems to enhance cyber security in telecommunication networks. *Scientific Bulletin of UNFU*, 34(6), 93–100. https://doi.org/10.36930/40340613

traditional, single-device training methods are often insufficient to handle the computational demands of modern ML algorithms, parallel processing through multiple GPUs provides an innovative way to address these shortcomings.

Scope and hypotheses: this research explores the hypothesis that multi-GPU setups can significantly reduce the time required to train machine learning models while improving their accuracy, particularly when MI-based batch size optimization is applied.

This article is structured to guide the reader through the scientific context and research process. Following this introduction, we present a comprehensive literature review, outlining existing research on the application of ML in cyber security and the use of multi-GPU setups for training. The subsequent methodology section details the experimental design, data collection, model training processes, and evaluation metrics used in this study. We then report our experimental results, comparing the performance of single-GPU and multi-GPU setups, with a focus on batch size optimization. Finally, we conclude by summarizing the key findings and their significance for future developments in cyber security ML models.

In conclusion, this research aims to deepen the understanding of how multi-GPU setups can enhance ML capabilities in the realm of cyber security. By building on prior research and foundational work in the field, this study contributes to the development of more efficient and resilient cyber security measures. The insights gained from the use of multi-GPU configurations, particularly in conjunction with batch size optimization strategies such as Mutual Information (MI), demonstrate their potential to protect telecommunication networks from an ever-evolving spectrum of cyber threats [4, 7, 11]. These advanced computational techniques enable more accurate and faster threat detection, a critical need in today's cyber security landscape [6, 8].

Object of research – the training of machine learning models designed for cyber security applications.

Subject of research – is the architecture of multi-GPU training and its effect on optimizing ML models, particularly through batch size determination and parallel data processing. This approach offers the opportunity to harness the full computational power of multiple GPUs, which will allow models to learn more efficiently, scale better with larger datasets, and deliver superior performance in real-time cyber security applications.

The purpose of the work – is to explore how multi-GPU setups can optimize the machine learning training process, leading to improved efficiency and accuracy for models used in cyber security. By achieving these improvements, telecommunication networks can better safeguard themselves from modern cyber threats, ensuring more robust and agile defense systems.

To achieve this *purpose*, the following main *research objectives are identified*:

1. Investigate the advantages of multi-GPU setups over single-GPU setups for ML model training, which will enable faster and more efficient training processes, ultimately allowing models to process larger datasets in less time. This improvement is crucial for maintaining up-to-date defenses against rapidly evolving cyber threats.
2. Examine the role of batch size optimization and the application of Mutual Information (MI) in determining optimal batch sizes, which will provide an opportunity to enhance model training by ensuring that batch sizes are selected in a way that maximizes learning efficiency while minimizing computational waste. The ability to fine-tune this parameter will result in more precise models with less time spent in training.
3. Evaluate the efficacy of manual versus MI-determined batch sizes in improving training speed and accuracy, which will ensure that the most effective batch sizing strategy is employed, thereby improving both the speed of training and the resulting model's ability to generalize well to unseen data. This, in turn, will help telecommunication networks respond more accurately and promptly to cyber-attacks.
4. Analyze the implications of multi-GPU training for real-time cyber security systems, which will help identify how these advancements can be practically implemented to improve the speed and reliability of cyber security frameworks. Integrating multi-GPU configurations will enhance the adaptability and robustness of these systems.
5. Provide recommendations for the future development of more resilient ML models in telecommunication networks, which will lead to improved defense strategies capable of addressing increasingly sophisticated cyber threats. By incorporating lessons learned from this research, future models can be designed to be even more efficient, scalable, and responsive.

Analysis of recent research and publications. The field of cyber security has undergone significant transformation due to the integration of machine learning (ML) technologies. ML has revolutionized threat detection, anomaly identification, and automated response systems, making cyber security frameworks more adaptive and dynamic. Unlike traditional rule-based systems that are inflexible and reactive, ML offers a system that continuously learns from data and adapts to evolving threats, enhancing the robustness of security measures.

Several recent studies have explored the applications of ML in cyber security. The authors of the study [1] provided an extensive review of machine learning techniques and their role in mitigating cyber security threats, emphasizing that the adaptability of ML systems allows for better threat detection and mitigation strategies. ML's ability to generalize across diverse threat landscapes is what gives it an edge over conventional approaches. Similarly, researchers [19] explored the role of ML in cyber security data science, pointing out that machine learning models can dynamically adapt to new forms of cyber-attacks and improve detection rates in real-time scenarios.

The practical applications of ML in cyber security are numerous, and one of the most significant is image recognition. Image recognition is increasingly being used to detect phishing attacks, where malicious actors use misleading images or icons to trick users into compromising sensitive information. It has been shown [18] that small and medium enterprises, in particular, have benefited from adopting ML-based image recognition systems to detect these sophisticated attacks. Additionally, the ImageNet Object Localization Challenge, a benchmark in image classification, has demonstrated how machine learning can drastically reduce image recognition errors, making it an essential tool for cyber security applications [10].

While the benefits of ML in cyber security are well-documented, the computational demands of training these models, particularly in image recognition, are substantial. The introduction of parallel processing, which has evolved to include multi-GPU setups, allows several GPUs to work concurrently, thus expediting the training process [16]. Multi-GPU systems effectively distribute computational

workloads across several devices, accelerating training times while improving performance [14].

It has also demonstrated that GPU-accelerated machine learning inference is crucial for real-time applications. In cyber security, where threats evolve rapidly, reducing training times without sacrificing model accuracy is critical. Multi-GPU setups offer a solution by providing near-linear scaling in performance, which is crucial for real-time security applications [21].

Despite the advantages of multi-GPU setups, a key challenge remains in optimizing batch sizes during training. Batch size optimization plays a critical role in determining the speed and quality of training. Larger batch sizes, as explored by Bharadiya (2023), often result in faster training but may compromise the model's ability to generalize, leading to lower validation accuracy. On the other hand, smaller batch sizes can improve accuracy but significantly prolong the training process [4].

Mutual Information (MI) has recently emerged as a promising technique for addressing this challenge. Research [20] explored the use of MI to enhance the training process in machine learning models for cyber security applications. MI quantifies the shared information between training data and model predictions, offering a more principled approach to selecting batch sizes. However, despite its promise, the use of MI in batch size optimization remains underexplored, particularly in the context of multi-GPU environments.

The scientific literature provides a solid foundation for understanding the computational benefits of multi-GPU setups in ML model training. However, critical gaps remain, particularly concerning batch size optimization. While other studies [11] have explored multi-GPU setups, batch size optimization has often been overlooked in the broader context of real-time cyber security applications.

Additionally, the integration of MI into batch size optimization has been explored only minimally. A recent study [3] highlighted the potential of MI for improving the accuracy of ML models, but their research did not fully delve into its applicability in multi-GPU setups. Addressing these gaps could lead to more efficient and accurate models, significantly enhancing cyber security defenses.

In conclusion, while significant progress has been made in the use of ML for cyber security, particularly in image recognition and multi-GPU training setups, important research gaps remain. The lack of focus on batch size optimization and MI integration in multi-GPU environments represents an area where further investigation is needed. By addressing these gaps, this research aims to contribute to the development of more efficient, real-time cyber security solutions that are capable of handling the rapidly evolving threat landscape.

Materials and methods of research. This study is structured as a comparative analysis aimed at evaluating the performance differences between machine learning (ML) models trained on single-GPU and multi-GPU setups. The research specifically focuses on the impact of batch size determination strategies – manual tuning and optimization via Mutual Information (MI) – on training efficiency and validation accuracy. The goal is to identify the setup and strategy that not only enhance model training speed but also optimize model accuracy on unseen data.

A variety of GPU configurations were explored, ranging from single-GPU systems to complex multi-GPU architectures. This diversity allows for a detailed examination of how parallel processing influences the time required for model training and the resulting accuracy of the models.

The primary criteria for evaluating model performance were validation accuracy and training time. Validation accuracy ensures that the model generalizes well to new, unseen data, while training time assesses the efficiency of the training process. These metrics were chosen to balance the need for high-performing models with the practical constraints of computational resources.

The study utilized datasets compiled from publicly available cyber security threat data. This data included a wide range of cyber threats, aiming to provide a robust dataset for comprehensive model training and validation [1, 10].

Prior to training, the data underwent standard preprocessing steps such as normalization and feature scaling. Data augmentation techniques were also applied to increase the dataset's diversity and volume, thereby enhancing model resilience against overfitting and improving its ability to generalize [19].

The research deployed various ML models, with a focus on those demonstrating high efficacy in similar cyber security tasks. The selection included, but was not limited to, convolutional neural networks (CNNs) and recurrent neural networks (RNNs), tailored to the specific characteristics of the cyber security data being analyzed [13, 16].

The training process highlighted two approaches to batch size determination [22]:

1. Manual Tuning: Batch sizes were initially set based on conventional practices and iteratively adjusted based on observed training performance and validation accuracy.
2. MI-Determined Batch Sizes: An innovative method was introduced wherein batch sizes were optimized based on the calculation of Mutual Information (MI) between training data batches, with a novel focus on maximizing validation accuracy and minimizing training time. The optimization aimed to find a batch size that balanced the trade-off between efficient training and high model accuracy on the validation set [7, 17].

The optimization goal was adjusted to consider both validation accuracy and training time, formulated as [14]:

$$[max]\_(batch\ size)\ Validation\ Accurance - \alpha \times Training\ Time$$

where α is a regularization parameter that balances the two objectives.

Model performance was assessed using validation accuracy and training time as primary metrics. Additional metrics included precision, recall, and F1 score to provide a comprehensive evaluation of model effectiveness in cyber security threat detection and classification [21].

This revised methodology ensures that the selection of the optimal batch size is directly aligned with practical outcomes, emphasizing not only the efficiency of the training process but also the effectiveness of the models in real-world cyber security applications.

Research results and their discussion / Результати дослідження та їх обговорення

Results. Our experimentation focused on comparing the effectiveness of single-device and multi-device training setups, with a particular emphasis on batch size optimization using Mutual Information (MI). The results from our tests are presented through a series of visual representations (Fig. 2-7) and comparative analyses.

Single-device Training. Setup and Execution: The single-device training served as our baseline, utilizing 1,024,977 images for training and 256,190 images for vali-

dation. The MobileNetV2 architecture was employed, running on a single GPU.

Results and Analysis: As shown in Figure 1, the training process on a single GPU exhibited a steep increase in training accuracy, but the validation accuracy plateaued early, indicating potential overfitting. The training loss decreased over time, while validation loss increased slightly, further reinforcing overfitting concerns. The training time totaled 106,878.89 seconds, a considerable duration compared to multi-device setups.

Multi-device Training with Manual Batch Sizing. *Setup and Execution:* Next, we shifted to multi-GPU training with manual batch size tuning. Different batch sizes were tested to analyze their impact on training efficiency and accuracy.

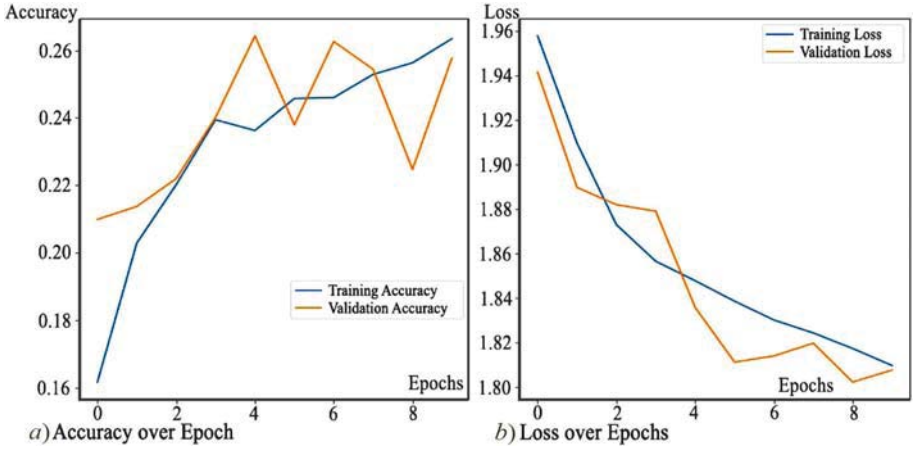

Figure 1. Training and validation accuracy and loss over epochs for single-device training / Точність тренування та валідації і втрати впродовж епох для тренування на одному пристрої

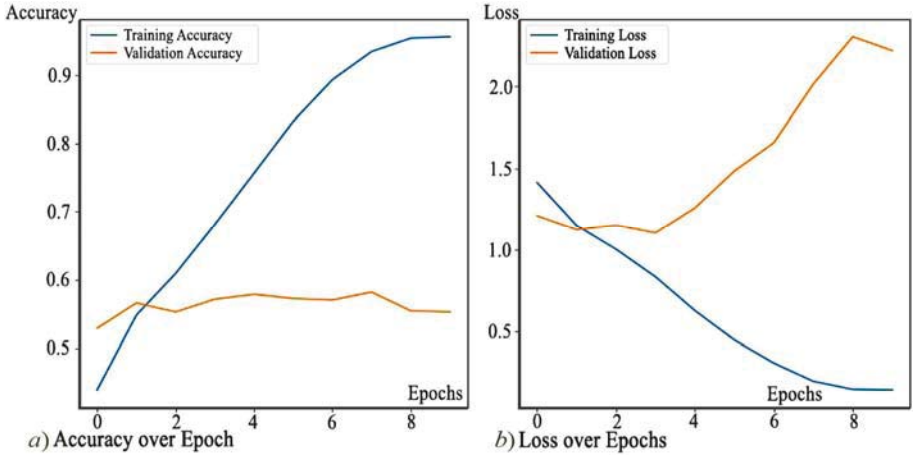

Figure 2. Training and validation accuracy and loss over epochs with manual batch sizing on multi-device setup / Точність тренування та валідації і втрати впродовж епох з ручним встановленням розміру пакету на багатопристроєвій конфігурації

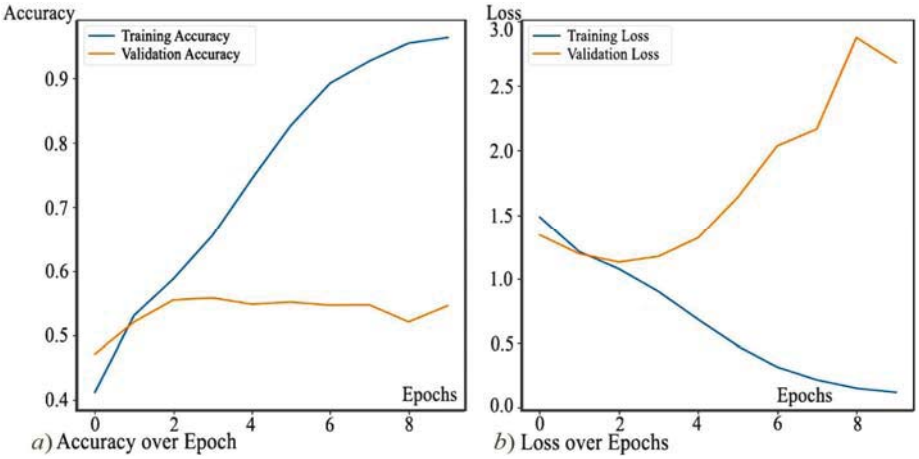

Figure 3. Training and validation accuracy and loss over epochs with manual batch sizing (Batch Size = 16) on multi-device setup / Точність тренування та валідації і втрати впродовж епох з ручним встановленням розміру пакету (Розмір пакету = 16) на багатопристроєвій конфігурації

Results and Analysis: The results shown in Figure 2 indicate that larger batch sizes reduced training time but at the cost of lower validation accuracy. These findings suggest a trade-off between speed and performance, commonly observed in batch size optimization.

Multi-device Training with MI-determined Batch Sizes. *MI Optimization Approach:* The MI-guided batch sizing aimed to optimize both training efficiency and model accuracy. MI was employed to determine batch sizes that balance computational load and model performance, hypot-

hesized to yield superior outcomes compared to manual batch sizing.

***Results and Analysis*:**

Batch Size = 16: as depicted in Figure 3, the MI-determined batch size of 16 produced a balance between training and validation accuracy. However, the training time was longer compared to larger batch sizes.

Batch Size = 32: figure 4 shows that increasing the batch size to 32 led to faster training times while maintaining reasonable accuracy. This configuration displayed better performance compared to the single-device baseline.

Batch Size = 64: the optimal performance was observed with a batch size of 64. As Figure 5 illustrates, this batch size provided the highest validation accuracy of 59.34 % and a significant reduction in training time to 45,392.73 seconds.

Batch Size = 128: Further increasing the batch size to 128 resulted in diminished returns, as seen in Figure 6. The validation accuracy dropped, confirming that excessively large batch sizes negatively impact performance, despite reducing training time.

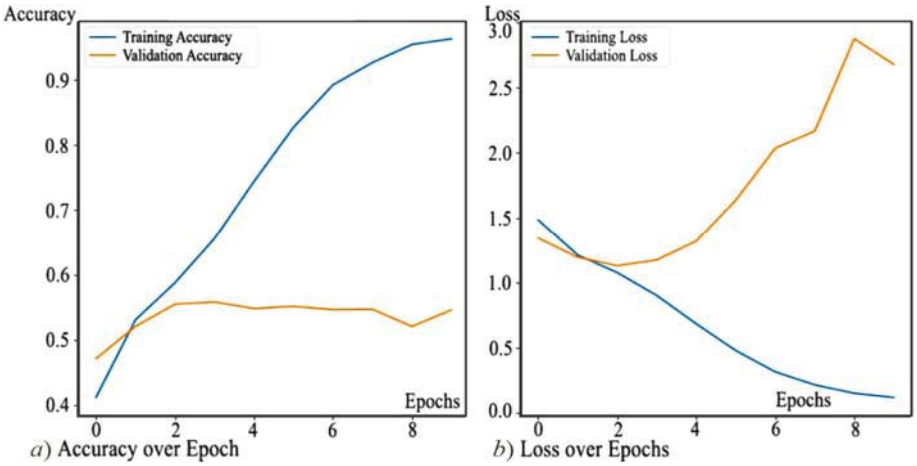

Figure 4. Training dynamics depicting accuracy and loss for a multi-GPU setup using manual batch sizing (Batch Size = 32) / Динаміка тренування, що показує точність та втрати для багато-GPU-конфігурації з використанням ручного встановлення розміру пакету (Розмір пакету = 32)

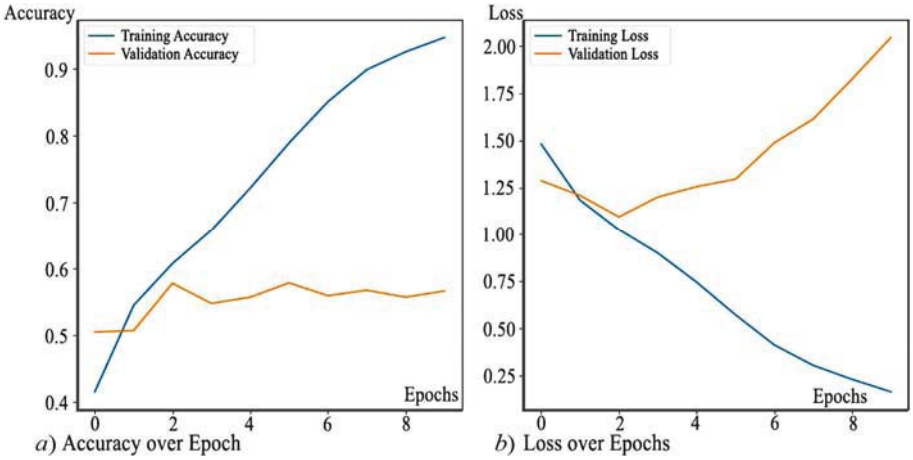

Figure 5. Optimization results of training with a batch size of 64 on a multi-GPU setup, showcasing the effectiveness of MI-determined batch sizing / Результати оптимізації тренування з розміром пакету 64 на багато-GPU-конфігурації, що демонструє ефективність встановлення розміру пакету на підставі MI

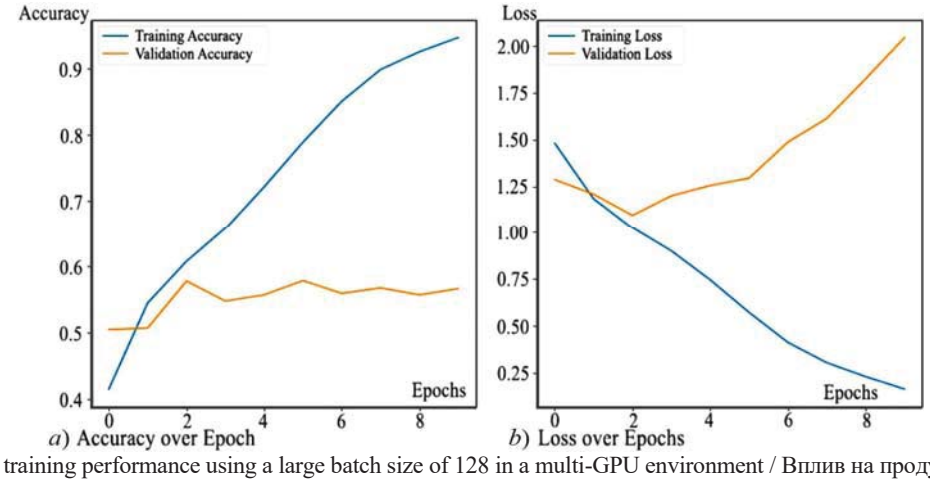

Figure 6. Impact on training performance using a large batch size of 128 in a multi-GPU environment / Вплив на продуктивність тренування з великим розміром пакету 128 у багато-GPU-середовищі

The final results confirm that MI optimization effectively balances training efficiency and model accuracy, making it a practical approach for machine learning model training in cyber security. The best performance was achieved with a batch size of 64, highlighting the critical role of strategic batch size optimization in multi-GPU environments.

Comparative analysis of research results. The Multi-GPU Neural Network Configuration, shown in Fig. 7, illustrates how data parallelism works when employing multiple GPUs. Each GPU handles different batches of the dataset (Batch Set 0 and Batch Set 1) simultaneously, optimizing the training process by distributing the computational load. This parallelism enables faster training times and improved overall performance, which is critical for real-time cyber security environments that demand swift model deployment.

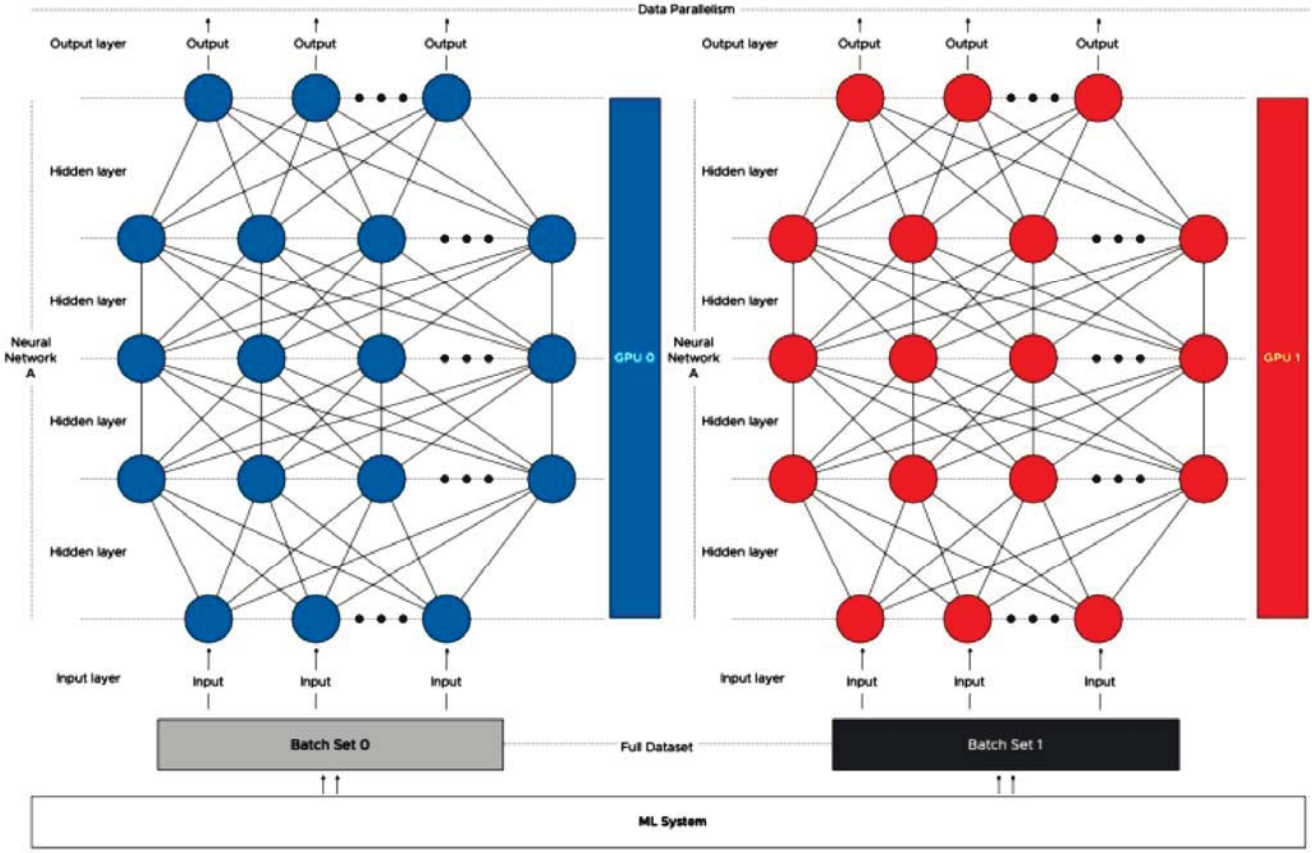

Figure 7. Multi-GPU Neural Network Configuration illustrating data parallelism, where two GPUs handle different batches of the dataset simultaneously to improve training efficiency and performance / Конфігурація нейронної мережі з декількома графічними процесорами, що ілюструє паралелізм даних, коли два графічні процесори одночасно обробляють різні партії даних для підвищення ефективності навчання та продуктивності

This technique of data parallelism, as depicted in Fig. 7, demonstrates how multiple GPUs can work concurrently to process different parts of a dataset. By allowing GPUs to work in parallel, we can significantly reduce training times and ensure that models are deployed more efficiently in cyber security environments where real-time responsiveness is essential.

Batch size optimization plays a crucial role in determining both the training time and accuracy of machine learning models. Larger batch sizes often lead to faster training times but can result in lower model accuracy, as they tend to dilute the information in each training iteration. Conversely, smaller batch sizes, while more precise, often require significantly longer training times. Our study explored the trade-offs between these two approaches, emphasizing the role of Mutual Information (MI) in optimizing batch sizes.

The MI-guided approach introduced a more nuanced optimization strategy. MI quantifies the predictive power of each batch, ensuring that every iteration is informational dense. By adjusting the batch size based on MI, our methodology succeeded in optimizing both training speed and accuracy, as demonstrated by the performance with batch size 64 [8, 20].

This process is mathematically represented by the principle that the estimation error is inversely proportional to the square root of the batch size, as shown in the following formula:

$$batchsize \in \frac{1}{\sqrt{batchsize}}. \quad (1)$$

This relationship suggests that there exists an optimal batch size that minimizes the error without causing significant computational overhead, striking a balance between the speed of convergence and model precision. As depicted in Fig. 4 and 6, increasing the batch size up to 64 significantly improved validation accuracy while reducing training time.

Mutual Information (MI) has emerged as a pivotal factor in enhancing the training process. By employing MI to determine batch sizes, our methodology aligns the training process more closely with the intrinsic data structure. MI effectively quantifies the shared information between the input data (X) and the target variables (Y), ensuring that each batch is informational dense. The mathematical formulation of MI is expressed as:

$$I(X;Y) = \sum_{x\in X, y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \quad (2)$$

This formula quantifies the amount of shared information between the input X and the output Y, guiding the optimization process to select batch sizes that are neither too small to cause noisy updates nor too large to dilute the informational content of each iteration.

By aligning the batch size with the data structure's mutual information, our approach ensures that the training process remains efficient while maximizing model accuracy.

Discussion of research results. This study's comparative analysis highlights key advancements in machine learning model training for cyber security applications, particularly in terms of efficiency and accuracy. We specifically analyzed the impact of single-device versus multi-device training setups, along with the effectiveness of batch size optimization using Mutual Information (MI).

The experiments clearly show that multi-device training setups outperform their single-device counterparts. As depicted in Fig. 2, 4, and 5, multi-GPU setups resulted in a more rapid descent in loss metrics and an accelerated increase in accuracy over fewer epochs. This improvement can be attributed to the reduction in time complexity offered by parallel processing. Multi-GPU training scales efficiently with the number of devices, providing near-linear gains in performance [11, 21].

The optimization of batch size played a crucial role in maximizing the computational benefits of multi-device setups. As previous studies in distributed computing suggest, optimizing the batch size improves both training efficiency and resource utilization [2, 8, 17]. The MI-determined batch sizes utilized in our study validate this claim, as they managed to balance the computational load while maintaining or enhancing model accuracy.

While larger batch sizes significantly reduced training times, they also introduced challenges. As shown in Fig. 6, increasing the batch size to 128 diminished validation accuracy, reaffirming that overly large batch sizes can dilute the informational content of each iteration, leading to sub-optimal model updates [3, 4]. This aligns with existing literature, which highlights the trade-off between batch size and model generalization capabilities [7, 20].

So, based on the results of the work performed, it is possible to formulate the following scientific novelty and practical significance of the research results.

Scientific novelty of the obtained research results – developed an advanced methodology for optimizing batch sizes in multi-GPU machine learning environments, using Mutual Information (MI) as a guiding metric, which balances computational efficiency and data structure alignment, ensuring enhanced training performance and improved model accuracy for real-time cyber security applications in telecommunication networks.

Practical significance of the research results – are applied in the development of more efficient machine learning models for cyber security in telecommunication networks, where rapid threat detection and adaptation are crucial. The MI-based batch size optimization strategy is implemented to significantly reduce model training time while maintaining or improving validation accuracy, ensuring that these models can be more effectively deployed in real-time cyber security applications. Additionally, the proposed method can be adapted and applied to other domains, such as the Internet of Things (IoT) and autonomous systems, where computational efficiency and model precision are equally important. The research findings provide a foundation for further advancements in multi-GPU architectures, with practical implications for industries requiring scalable, high-performance machine learning solutions.

Conclusions / Висновки

Peculiarities of setting up multi-GPUs have been studied in terms of the possibility of optimizing the process of machine learning training process, leading to improved efficiency and accuracy for models used in cyber security. By achieving these improvements, telecommunication networks can better safeguard themselves from modern cyber threats, ensuring more robust and agile defense systems. Based on the results of the research the following main conclusions can be drawn.

1. Established that multi-GPU configurations significantly enhance the training efficiency and accuracy of ML models, particularly for real-time cyber security tasks in telecommunication networks.
2. Developed an advanced MI-guided approach for optimizing batch sizes, which improves both computational efficiency and model accuracy, outperforming traditional manual tuning methods in large-scale datasets.
3. Found that the use of MI-based batch size optimization achieves a better balance between training speed and model accuracy, particularly with a batch size of 64, which proved optimal in reducing training time while maintaining high validation accuracy.
4. Investigated the trade-off between batch size and model performance, confirming that excessively large batch sizes (e.g., 128) reduce validation accuracy, whereas MI-optimized sizes mitigate such performance degradation.
5. Presented experimental evidence showing that multi-GPU setups, when combined with MI-based batch size optimization, enable faster model training, which is critical for the rapid deployment of cyber security solutions.
6. Analysed the impact of MI-guided optimization on reducing the risk of overfitting, which is essential for improving the generalization capabilities of ML models in detecting cyber security threats.
7. Introduced the concept of applying MI-driven optimization strategies to real-time cyber security frameworks, providing a more agile and effective defence mechanism against evolving cyber threats.
8. Suggested future research into expanding the application of MI-based optimization in other domains such as the Internet of Things (IoT) and autonomous systems, where computational efficiency and accuracy are equally critical.
9. Applied this methodology to address the specific computational challenges in telecommunication network security, resulting in a scalable approach that enhances the overall cyber security posture.
10. Proposed further exploration of MI refinement techniques, distributed computing, and quantum ML as potential avenues for advancing the effectiveness of ML models in cybersecurity.

References

1. Ahsan, M., Nygard, K. E., & Cramer, G. R. (2022). Cyber security threats and their mitigation approaches using machine learning – a review. *Journal of Cyber security and Privacy*, 2(3), 527–555. https://doi.org/10.3390/jcp2030027
2. Aljuhani, A. (2021). Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments. In IEEE Access, Vol. 9, pp. 42236–42264. https://doi.org/10.1109/ACCESS.2021.3062909
3. Apruzzese, G., Laskov, P., Montes de Oca, E., Mallouli, W., Rapa, L. B., Grammatopoulos, A. V., & Di Franco, F. (2023). The role of machine learning in cyber security. *Digital Threats: Research and Practice*, 4(1), 1–38. https://doi.org/10.1145/3545574

4. Bharadiya, J. (2023). Machine learning in cyber security: Techniques and challenges. *European Journal of Technology*, 7(2), 1–14. https://doi.org/10.47672/ejt.1486
5. Choi, S., Lee, S., Kim, Y., Park, J., Kwon, Y., & Huh, J. (2021). Multi-model machine learning inference serving with GPU spatial partitioning. https://doi.org/10.48550/arXiv.2109.01611
6. Das, R., & Morris, T. H. (2017). Machine learning and cyber security. In: *Proceedings of the 2017 International Conference on Computer, Electrical & Communication Engineering* (*ICCECE*), 1–7. https://doi.org/10.1109/ICCECE.2017.8526232
7. Dasgupta, D., Akhtar, Z., & Sen, S. (2022). Machine learning in cyber security: A comprehensive survey. *The Journal of Defense Modeling and Simulation*, 19(1), 57–106. https://doi.org/10.1177/1548512920951275
8. Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using a deep learning approach for the Internet of Things. *Future Generation Computer Systems*, 82, 761–768. https://doi.org/10.1016/j.future.2017.08.043
9. Handa, A., Sharma, A., & Shukla, S. K. (2019). Machine learning in cyber security: A review. *WIREs Data Mining and Knowledge Discovery*, 9(4). https://doi.org/10.1002/widm.1306
10. Howard, A., & Park, E. (2018). ImageNet object localization challenge. *Kaggle*. URL: https://kaggle.com/competitions/imagenet-object-localization-challenge
11. Kozik, R., Choraś, M., Ficco, M., & Palmieri, F. (2018). A scalable distributed machine learning approach for attack detection in edge computing environments. *Journal of Parallel and Distributed Computing*, 121, 65–74. https://doi.org/10.1016/j.jpdc.2018.03.006
12. Medykovsky, M., Droniuk, I., Nazarkevich, M., & Fedevych, O. (2013). Modelling the perturbation of traffic based on ateb-functions. In: *Proceedings of the 20th International Conference on Computer Networks*, 38–44. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38865-1_5
13. Nazarkevych, M., Oliiarnyk, R., Nazarkevych, H., Kramarenko, O., & Onyshchenko, I. (2016). The method of encryption based on Ateb-functions. In: *Proceedings of the IEEE First International Conference on Data Stream Mining & Processing* (*DSMP*), 129–133. https://doi.org/10.1109/DSMP.2016.7583543
14. Pal, S., Mars, J., Subramoney, S., & Chrysos, G. (2019). Optimizing multi-GPU parallelization strategies for deep learning training. *IEEE Micro*, 39(5), 91–101. https://doi.org/10.1109/MM.2019.2935967
15. Pandey, M., Fernandez, M., Gentile, F., & Sheel, A. (2022). The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence*, 4, 211–221. https://doi.org/10.1038/s42256-022-00463-x
16. Parhami, B. (1999). Introduction to parallel processing. *Springer*. https://doi.org/10.1007/b116777
17. Rashid, Z. N., Zebari, S. R. M., & Jacksi, K. (2018). Distributed cloud computing and distributed parallel computing: A review. In: *Proceedings of the 2018 International Conference on Advanced Science and Engineering* (*ICOASE*), 167–172. https://doi.org/10.1109/ICOASE.2018.8548937
18. Rawindaran, N., Jayal, A., & P. E. (2021). Machine learning cyber security adoption in small and medium enterprises in developed countries. *Computers*, 10(11). https://doi.org/10.3390/computers10110150
19. Sarker, I. H., & Alam, K. B. S. (2020). Cyber security data science: An overview from a machine learning perspective. *Journal of Big Data*, 7(1), 55. https://doi.org/10.1186/s40537-020-00318-5
20. Sivanathan, A., Habibi Gharakheili, H., & Sivaraman, V. (2020). Managing IoT cyber-security using programmable telemetry and machine learning. *Transactions on Network and Service Management*, 17(1), 60–74. https://doi.org/10.1109/TNSM.2020.2971213
21. Wang, M., Yang, T., Flechas, M. A., Harris, P., Holzman, B., Knoepfel, K., & Krupa, J. (2021). GPU-accelerated machine learning inference as a service for computing in neutrino experiments. *Frontiers in Big Data*, 3. https://doi.org/10.3389/fdata.2020.604083
22. Wiley, J. (2023). Meta-heuristic algorithms for advanced distributed systems. Wiley. https://doi.org/10.1002/9781394188093

О. П. Кузів, М. А. Назаркевич

Національний університет "Львівська політехніка", м. Львів, Україна

ОПТИМІЗАЦІЯ ПРОЦЕДУРИ МАШИННОГО НАВЧАННЯ МОДЕЛІ НА БАГАТОПРОЦЕСОРНИХ УСТАНОВКАХ ДЛЯ ПОСИЛЕННЯ КІБЕРБЕЗПЕКИ В ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖАХ

Проаналізовано особливості оптимізації процедури машинного навчання ML (англ. *Machine Learning*) моделей із використанням багатопроцесорних графічних установок (multi-GPU) для підвищення кібербезпеки в телекомунікаційних мережах. Одним із ключових аспектів дослідження стало використання паралельного оброблення даних, яке дає змогу розподілити навчальне навантаження між кількома графічними процесорами, що значно скорочує тривалість навчання та підвищує точність моделей, що особливо важливо для швидкого виявлення загроз у кіберпросторі. Запропоновано новий підхід до оптимізації розміру партій даних (англ. *batch size*) за допомогою взаємної інформації MI (англ. *Mutual Information*), що дає змогу гармонізувати використання обчислювальних ресурсів і інформаційного вмісту навчальних даних. Використання MI допомагає визначати оптимальний розмір партій даних, який мінімізує похибки навчання та підвищує точність моделі без значного збільшення часу на навчання. Багатопроцесорні конфігурації, як показали результати експериментів, демонструють істотні переваги порівняно з однопроцесорними установками, забезпечуючи швидше навчання та покращені показники точності моделей. Особливо наголошено, що MI-кероване налаштування розміру партій даних значно перевершує традиційні методи ручного налаштування, забезпечуючи більш високу точність валідації та зменшуючи тривалість навчання. За результатами дослідження з'ясовано, що підхід на підставі взаємної інформації є ефективним інструментом для вирішення проблеми оптимізації процесу навчання моделей ML у реальних сценаріях, де кібербезпека є критичним елементом. Запропоновані методи дають змогу ML моделям швидше навчатися та точніше ідентифікувати потенційні загрози, що робить їх особливо актуальними для телекомунікаційних мереж, де необхідна швидка реакція на нові загрози в режимі реального часу. Запровадження сучасних обчислювальних технологій, таких як багатопроцесорні графічні установки та MI-оптимізоване навчання, підвищує ефективність і точність моделей машинного навчання. Це дає змогу поліпшити заходи з кібербезпеки і гарантує надійніший захист телекомунікаційних мереж від зловмисних атак. Зазначено, що запропоновані підходи можуть бути адаптовані не тільки для кібербезпеки, а й для інших галузей, де важлива висока точність моделей і швидке навчання. Перспективи подальших досліджень охоплюють розвиток нових методів машинного навчання, зокрема глибинних нейронних мереж, дослідження інших альтернативних обчислювальних архітектур, таких як квантові обчислення чи розподілені системи, а також їх інтеграція в реальні системи у режимі реального часу. Окрему увагу доцільно приділити етичним аспектам застосування автоматизованих систем кібербезпеки, зокрема, щодо уникнення упередженості в алгоритмах та забезпечення рівноправності під час їх використання.

Ключові слова: паралельне оброблення; оптимізація процедури навчання; виявлення кіберзагроз; машинне навчання в кібербезпеці; взаємна інформація.