Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

WEAPON COMBAT EFFECTIVENESS ANALYTICS: INITIAL INSIGHTS USING DEEP LEARNING AND BIG DATA FROM VC SIMULATIONS

Luis Rabelo[1], Larry Lowe[1], Won Jung[1], Marwen Elkamel[1], and Gene Lee[1]
[1]Industrial Engineering & Management Systems, University of Central Florida, Orlando, FL, USA

ABSTRACT

This paper presents initial insights into applying deep learning and big data analytics for assessing Weapon Combat Effectiveness (WCE) in dynamic combat environments. Traditional WCE models often rely on simplified assumptions and limited variables, restricting their real-world applicability. We address these limitations using datasets from integrated Virtual-Constructive (VC) simulation frameworks, combining strengths of defense modeling, big data, and AI. Our experiments focus on a case study with two opposing forces: Blue Force of seven F-16s, and Red Force of two surface-to-air missile units and a high-value facility. Using raw simulation data—filtered and enhanced to isolate key Measures of Performance —we trained a convolutional neural network to model nonlinear relationships and evaluate mission success probability. Initial findings demonstrate the model's robustness in handling data noise and its potential to support decision-making through visualizations. Early results suggest that deep learning, when integrated with federated VC simulations, can significantly enhance WCE analytics.

1 INTRODUCTION

Although live simulation may appear to be the most realistic and desirable form of training or experimentation, it presents several significant limitations when used as a standalone solution. Live simulation refers to training exercises conducted in real-world environments using actual personnel and physical assets—such as aircraft, ships, weapons, and troops—operating under simulated conditions (e.g., mock threats, scripted scenarios, or inert ordnance). These simulations aim to replicate operational situations as closely as possible but involve high resource consumption and logistical complexity. Several factors make live simulations less practical:

- Resource Constraints: Access to physical assets may be limited due to availability, maintenance schedules, or competing operational demands (Kirby et al. 2011).
- Lack of Reproducibility: Unlike digital simulations, live exercises are difficult to repeat under identical conditions. This can lead to variability in training experiences across units and hinder standardization of instruction (West et al. 2013).
- Cost and Safety Risks: Live simulations often entail significant financial costs and potential safety hazards. For example, the U.S. Department of Defense spent approximately one-third of its defense budget on training activities in 2012 (Summers 2012).
- Training Transfer Issues**:** In some cases, skills or behaviors reinforced during live simulations may not transfer effectively or even be detrimental when transitioning to real combat scenarios (Summers 2012).

This research aims to demonstrate the efficacy of Virtual and Constructive (VC) simulations in generating rich datasets to evaluate Weapon Combat Effectiveness (WCE). WCE refers to the quantitative and qualitative assessment of a weapon system's performance in achieving its intended tactical or strategic goals under varying operational conditions. For instance, assessing the WCE of a surface-to-air missile might involve analyzing its hit probability against different aircraft types in various weather and electronic warfare environments. The study employs collaborative simulation environments integrating virtual simulations (for real-time, human-in-the-loop scenarios) and constructive simulations (for automated, large-scale scenario generation) to facilitate this. These environments operate using Distributed Interactive Simulation (DIS) and High-Level Architecture (HLA) protocols (Lee et al. 2019), enabling interoperable and scalable simulation frameworks.

By conducting multiple scenario playthroughs—manual in virtual simulations and automated in constructive simulations—the system generates extensive datasets suitable for both performance-driven (e.g., missile range, hit/miss ratio, timing) and judgment-driven (e.g., operator decision quality, tactical adaptation) analysis. These simulations will log various outputs, including simulation logs, human surveys, and text messages, offering a comprehensive view of the scenario dynamics.

The collected data will then be processed using Machine Learning (ML) algorithms to detect patterns, identify key performance indicators, and develop predictive models of weapon performance. For example, by analyzing hundreds of engagements, ML can reveal under which conditions a weapon system's effectiveness drops, such as delays in operator response or environmental interferences.

We can create terabytes/petabytes of data with limited expert human interactions to guide the proposed system. Our focus will be on self-play. Tesauro (1995) suggests using "lesson-and-practice" training, which was successfully implemented by the HOYLE system. This method follows a steady pattern of expert human play, called the lesson, coupled with long self-play periods called the practice (Epstein 1994). For this research's scope, our learning system's training will come from virtual and constructive simulations. The utilization of AI (Rabelo et al. 2018; Rabelo et al. 2021; Cortes et al. 2020) has the potential to capture relationships and determine an acceptable WCE. The case study utilized is based on the seminal work of Jung (Jung 2018; Jung et al. 2017; Jung et al. 2019). Jung (2018) presented a sophisticated framework that includes Big Data and VC Simulations to determine WCE. This paper will expand on previous work where linear regression was utilized and instead leverage deep learning to capture relationships for WCE.

The remaining sections of the paper are organized as follows. Section 2 presents the case study employed in this paper, which utilizes VC simulation to model WCE. Section 3 details the computational model and the process for data generation. Section 4 outlines the deep learning framework that is trained on simulation outputs. Section 5 discusses the results, Section 6 concludes the paper, and discusses potential future work.

2 CASE STUDY

The case study implemented in this research builds upon the foundational work conducted by Jung (2018), which explored the use of Big Data and Virtual-Constructive (VC) simulations for modeling WCE. In the scenario, two opposing forces—Blue and Red—engage in a simulated combat operation. The Blue force comprises a squadron of seven F-16 fighter aircraft, tasked with executing an airstrike. The Red force is defended by two SA-8 TELAR surface-to-air missile (SAM) systems and a strategically significant core facility, which serves as the primary target of the Blue team. The mission objective for the Blue force is to successfully neutralize the Red team's core facility while minimizing aircraft losses, thus simulating a real-world operational tradeoff between mission success and force survivability.

Jung's original framework proposed leveraging a federated VC simulation architecture to collect large-scale data across numerous combat scenarios, which could then be used to model WCE through conventional statistical techniques. While this approach marked a significant step forward in simulation-based analytics, our research seeks to extend and significantly enhance it by incorporating deep learning methodologies. Specifically, we hypothesize that deep learning offers powerful capabilities for addressing the key analytical challenges inherent in WCE modeling—namely, the presence of noisy data, the need for extrapolation, the complexity arising from high-dimensional data spaces, and the nonlinear relationships among key operational parameters.

Deep learning models, particularly when architected with sufficient depth and trained on extensive scenario data, can discover latent patterns and decision rules that traditional regression-based methods are ill-equipped to capture. Additionally, deep learning models are inherently well-suited for handling the "Variety" dimension of Big Data. This includes integrating structured data—such as numerical inputs from telemetry logs—and unstructured data—such as textual battlefield reports, audio recordings from mission communications, and visual data from cockpit video feeds. By accommodating this heterogeneous data landscape, deep learning allows for a more holistic and nuanced analysis of combat effectiveness, reflecting quantitative performance metrics and qualitative expert assessments. Ultimately, this enhanced case study seeks to replicate the analytical framework developed in Jung's model and to significantly advance it by integrating scalable, intelligent systems capable of learning from high-dimensional, noisy, and heterogeneous datasets. This transformation enables the development of next-generation decision-support tools for Weapon Combat Effectiveness (WCE), grounded in robust data analytics and artificial intelligence. In this scenario, the Blue team's mission success is determined by its ability to destroy the Red team's core facility while minimizing its casualties—a classic tradeoff between offensive effectiveness and operational survivability. A comprehensive sensitivity analysis was conducted to assess performance across a wide range of combat conditions. The hundreds of simulations scenarios were run five times to capture stochastic variation. Each simulation ran 345 seconds, offering a rich dataset for subsequent analysis.

To quantitatively assess the outcomes of these simulations, the evaluation framework employed two key metrics commonly used in defense systems analysis:

- Measure of Effectiveness (MOE): A Measure of Effectiveness represents a high-level outcome metric that reflects a mission's operational success. In this study, the selected MOE was the Blue team's survival rate, as determined by the number of F-16 aircraft that remained operational at the conclusion of each mission scenario. This measure captures the overall effectiveness of the mission from a survivability perspective and aligns with strategic objectives such as force preservation and mission continuity.
- Measures of Performance (MOPs): Measures of Performance are quantitative indicators that reflect the technical and tactical performance of specific system components or subsystems during mission execution. These are lower-level indicators that influence the MOE. Seventeen MOPs were used in the case study, capturing various operational attributes of the Blue and Red teams. Specifically, two MOPs were defined for the Blue team—the number of bombs released and the quantity of chaffs/flares deployed—while five MOPs were assigned to the Red team, including variables like fire control range, tracking accuracy, missile inventory, and response timing. These MOPs serve as critical inputs for modeling and understanding how tactical choices and system configurations affect mission-level outcomes.

By combining these structured metrics with advanced deep learning techniques, the enhanced model provides accurate predictive analytics and deeper insights into the causal relationships between tactical inputs (MOPs) and strategic outcomes (MOE). This capability is essential for optimizing mission planning, weapon configuration, and operator training in future combat scenarios:

1. Blue Team: The number of released bombs, with a potential integer range of $1 \leq X_1 \leq 7$.
2. Blue Team: The number of chaffs and flares, with a potential integer range of $10 \leq X_2 \leq 100$.
3. Red Team: The control fire range, with a potential float range of $1 \leq X_3 \leq 10$.
4. Red Team: The control track range, with a potential float range of $1 \leq X_4 \leq 20$.
5. Red Team: The difference between control fire and track range, with a potential float range of $0 \leq X_5 \leq 19$.
6. Red Team: The control time between fires, with a potential float range of $5 \leq X_6 \leq 60$.
7. Red Team: The number of missiles, with a potential integer range of $0 \leq X_7 \leq 6$.

The training profile consists of two MOPs, one for each team:

1. Blue Team: The observation error, with a potential float range of $0 \leq X_8 \leq 100$.
2. Red Team: The reaction time, with a potential float range of $1 \leq X_9 \leq 100$.

The operations plan consists of four MOPs for the Blue team and four MOPs for the Red team:

1. Blue Team: Attack method 1, with a potential binary value of $X_{10} = (0,0)$.
2. Blue Team: Attack method 2, with a potential binary value of $X_{10} = (0,1)$.
3. Blue Team: Attack method 3, with a potential binary value of $X_{10} = (1,0)$.
4. Blue Team: The number of attack flights, with a potential integer range of $1 \leq X_{11} \leq 7$.
5. Red Team: Deployment 1, with a potential binary value of $X_{12} = (0,0)$.
6. Red Team: Deployment 2, with a potential binary value of $X_{12} = (0,1)$.
7. Red Team: Deployment 3, with a potential binary value of $X_{12} = (1,0)$.
8. Red Team: The number of defense SAMs, with a potential integer range of $1 \leq X_{13}$

3 COMPUTATIONAL MODEL

The case study was developed using a federated simulation architecture that integrates SIMbox and VR-Forces to support virtual and constructive simulations. This environment builds on enhancements described by Kim et al. (2014a and 2014b) and enables the modeling of realistic air-to-ground combat scenarios involving dynamic interactions between opposing forces. The complete system architecture is illustrated in Figure 1.

The Blue Team comprises multiple F-16 platforms, modeled using SIMbox simulators and constructive agents. These entities are built with source/user-defined C++ code, integrated with MÄK RTI (Run-Time Infrastructure), and utilize the DIS (Distributed Interactive Simulation) protocol for interoperability. Each F-16 simulation node represents a unit operating in a networked environment that mirrors real-world mission execution.

The Red Team includes several ground-based constructive components:

- Surface-to-Air Missile (SAM) systems modeled in SIMbox,
- A core facility representing the mission's primary target, and
- A support system developed in VR-Forces incorporates a data logger and HLA (High-Level Architecture) interface for mission command and support operations.

All simulation elements communicate through a DIS/HLA interoperability layer, which allows real-time interaction and synchronization of the federated entities during each scenario run.

Simulation outputs are recorded and stored in the Hadoop Distributed File System (HDFS) to support scalable data management across large scenario batches. Following each run, outputs are validated by a team of Subject Matter Experts (SMEs) to ensure fidelity and operational relevance. The validated data is then processed using a MapReduce framework, where:

- The MapReduce key is defined as the Blue Team's mission success, specifically destroying the core facility.
- The MapReduce values include the F-16 survival rate and associated Measures of Performance (MOPs) recorded for each simulation instance.

This federated environment enables the generation of statistical insights, such as probability maps and mission effectiveness evaluations, thereby supporting data-driven decision-making for combat mission planning and strategy refinement.

A comprehensive statistical analysis was conducted following data normalization and the initial selection of Measures of Performance (MOPs). This included checks for multicollinearity, heteroscedasticity, and outliers, which were addressed using appropriate methods such as deleting variables or scenarios, Cook's distance, and robust regression techniques. Heteroscedasticity was ultimately found to be absent, indicating consistent variance across the dataset and supporting the reliability of subsequent modeling steps.

To identify the most predictive variables for the Weapon Combat Effectiveness Equation (WCEE), a stepwise selection using the Akaike Information Criterion (AIC) was employed. This resulted in a final set of ten MOPs: X_1, X_2, X_3, X_7, X_8, X_9, X_{10b}, X_{11}, X_{12a}, and X_{12b}. When assigned specific values, these inputs yielded a high Measure of Effectiveness (YMOE = 95.90). Statistical validation using t-tests and ANOVA confirmed the feasibility of these WCEE models. Further evaluation using MMRE and PRED supported model accuracy, enabling the comparison—and, when needed, ensemble development—of WCEEs. From here, the optimal MOE and MOPs were extracted: YMOE = 95.90, $X_1 = 5$, $X_2 = 10$, $X_3 = 5.00$, $X_7 = 3$, $X_8 = 10.00$, $X_9 = 10.00$, $X_{10b} = 1$, $X_{11} = 2$, $X_{12a} = 0$, and $X_{12b} = 0$.

During the attribute analysis phase, MOP5 (the difference between fire control and tracking ranges) was identified as a derived variable mathematically dependent on MOP3 and MOP4. While such derived metrics can introduce multicollinearity and violate assumptions of predictor independence, we deliberately included MOP5 in the exploration phase to evaluate its marginal contribution to prediction accuracy. This allowed the model to autonomously assess and eliminate redundant features during the Akaike Information Criterion (AIC)-based stepwise selection process. As expected, MOP5 was ultimately excluded from the final Weapon Combat Effectiveness Equation (WCEE), which aligned with standard practices in regression modeling and reinforced the robustness of the variable selection process.

The final effectiveness values are presented as a probability map, which is meant to act as a visual aid for the command team to scan for optimal areas quickly. An example is shown in Figure 2. The map has two axes representing different Blue and Red forces' MOP values.

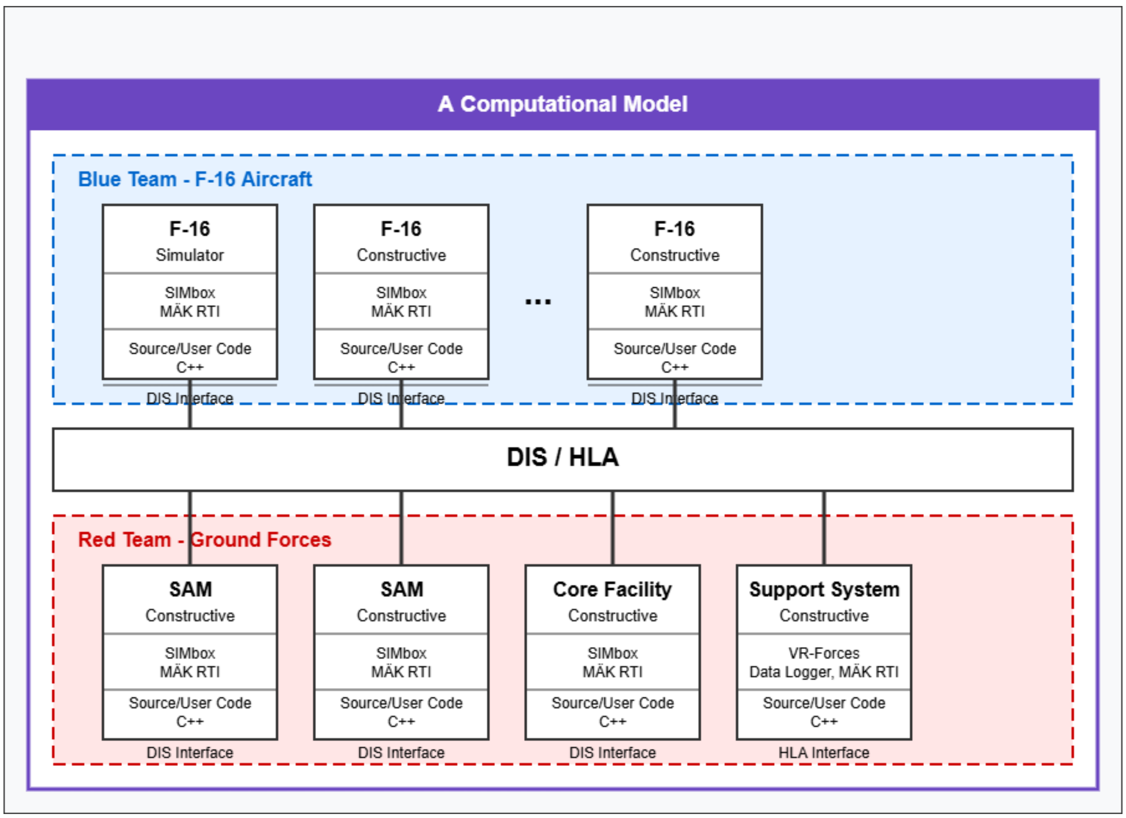

Figure 1: Computational model for virtual constructive simulation.

The probability map serves as a decision-support tool for the command team, enabling rapid identification of effective operational configurations. The horizontal axis represents selected Blue Team MOPs (e.g., X_1, X_3, ..., X_n), while the vertical axis represents varying Red Team conditions (e.g., X_2, X_4, ..., X_m). Each cell on the map indicates the computed probability of mission success for a given combination of force attributes, with values ranging from low to high.

High-probability zones (e.g., ≥90%) highlight favorable configurations for Blue Team operations, guiding tactical adjustments and force deployment strategies. Conversely, low-probability regions signal operational vulnerabilities under specific Red Team responses or environmental conditions. This visual representation allows planners to intuitively assess performance trade-offs and optimize mission planning under complex, data-driven conditions.

			Blue Team Factors									
			X_1		X_3				⋯	X_n		
			0	1	0.1	0.5	0.9	1.3	⋯	1.2	⋯	4.8
Red Team Factors	X_2	0.2	48%	99%	13%	18%	23%	50%	⋯	100%	⋯	38%
		0.6	13%	18%	53%	59%	71%	99%	⋯	80%	⋯	21%
		1.0	0%	15%	99%	81%	76%	55%	⋯	75%	⋯	11%
	X_4	12	3%	67%	54%	64%	74%	87%	⋯	76%	⋯	32%
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋯	⋮
		28	34%	78%	59%	73%	79%	92%	⋯	94%	⋯	53%
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋯	⋮
	X_m	1.4	15%	19%	57%	61%	73%	98%	⋯	82%	⋯	28%
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋯	⋮
		3.6	5%	69%	54%	68%	84%	97%	⋯	86%	⋯	42%

Figure 2: Example of MOE probability map based on MOPs influencing operations plan.

4 UTILIZATION OF DEEP LEARNING

Previous research on Weapon Combat Effectiveness (WCE), particularly the work by Jung (2018), laid the groundwork for modeling mission outcomes using structured data from Virtual-Constructive (VC) simulations and linear regression techniques. However, this approach presents several limitations that constrain the development of more intelligent and robust analytics. Specifically, four critical gaps were identified: (1) the exclusive reliance on structured data, omitting valuable unstructured sources such as audio, video, and mission communications; (2) the absence of methods for handling data noise, which is a core challenge in real-world, large-scale simulations; (3) the use of linear regression, which assumes linearity and leads to the exclusion of nonlinear but potentially important variables such as X_{13} (number of defense SAMs); and (4) the dependence on MapReduce, which in this case hindered flexible and efficient data exploration.

This study presents initial experiments that overcome prior limitations by applying deep learning to raw big data generated from federated Virtual-Constructive (VC) simulations. Over 314,000 simulation entries were filtered to extract the most relevant Measures of Performance (MOPs), reducing dimensionality and improving training efficiency. MapReduce was excluded to enable more agile data processing, and controlled noise was added to test model robustness. The refined dataset was used to train a convolutional neural network (CNN) using MOPs as inputs. Literature searches in Scopus and other databases revealed no existing studies combining deep learning, WCE, big data, and VC simulations—underscoring the novelty of this work.

In advancing beyond traditional regression techniques, CNNs were selected for their superior ability to manage high-dimensional, nonlinear, and noisy data characteristics intrinsic to WCE analytics. Unlike support vector machine (SVM) regression and boosting algorithms such as XGBoost, CNNs can automatically detect latent patterns, nonlinear dependencies, and feature interactions without requiring extensive manual preprocessing or feature engineering. In this study, CNNs generalized the heterogeneous data landscape well and captured local and global relationships among MOPs.

While SVMs and XGBoost are widely regarded for their strong performance in structured tabular data, they often depend on meticulous hyperparameter tuning and risk underfitting or overfitting when key interactions are not explicitly modeled. By contrast, our CNN architecture—built and tested in the KNIME environment—achieved high predictive accuracy with minimal intervention. The final configuration, which used optimized learning rates, batch sizes, and Xavier-based weight initialization, achieved an average prediction error of just 0.11%. This result significantly outperformed previous models, including Jung's linear regression-based approach, validating the choice of deep learning as the next step in WCE analytics.

The KNIME environment was utilized to build the flow of the initial experiment using deep learning convolutional neural networks (Ibrahim and Rabelo 2021). KNIME is a data analytics platform that uses various modules to implement Machine Learning and Data Mining solutions, while requiring no programming from the user (and using R and Python can be enhanced). This usability is achieved through an intuitive workflow-based user interface. In addition, KNIME uses libraries to implement dozens of Machine Learning algorithms. Figure 3 shows the flow built for this experiment.

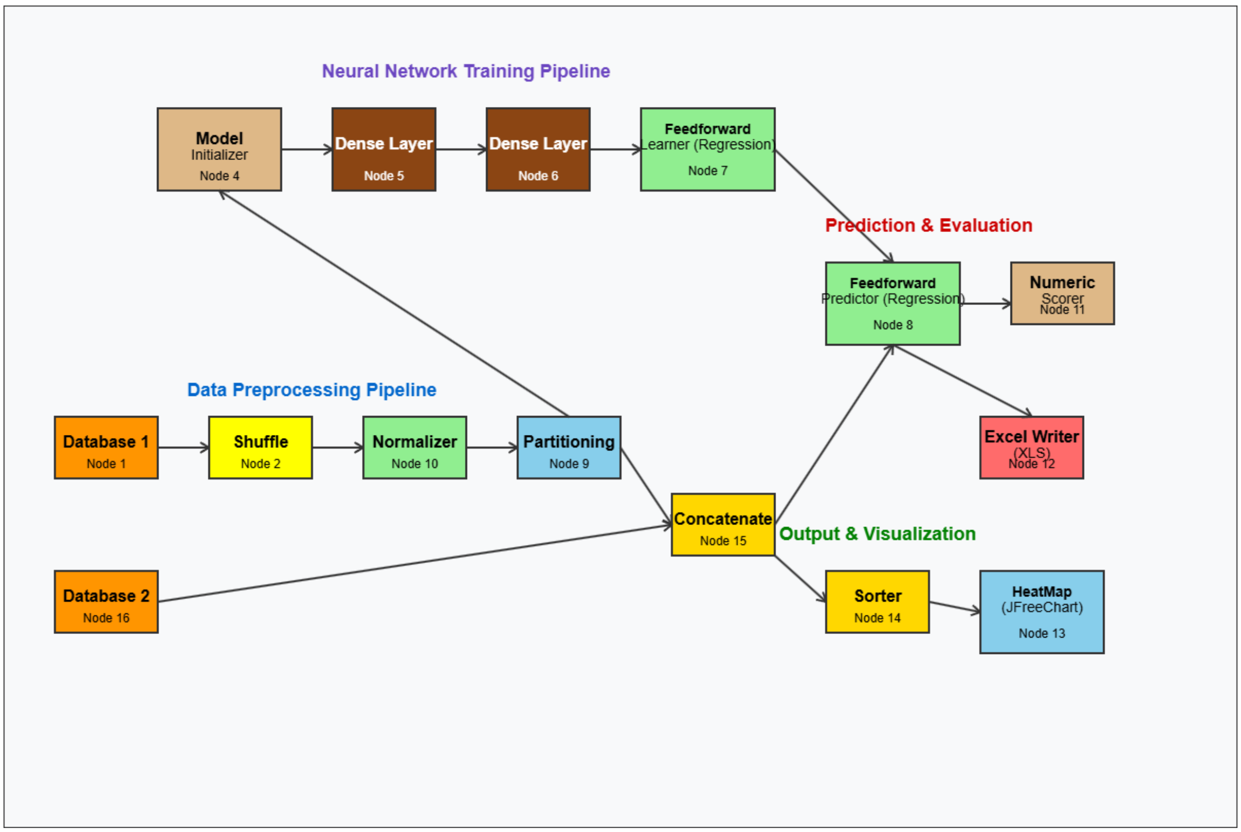

Figure 3: Example of deep learning implementation.

Several decisions and numerous trials were attempted to produce a deep learning configuration with the highest accuracy:

- **Architecture**: The architecture selected consisted of four layers (one input layer, two dense layers, and one output layer).
- **First Dense Layer Features**: The first dense layer uses ten neurons, a learning rate of 0.1, an activation function of the sigmoid (good for regression problems), and a Xavier type of initialization of weights. Xavier's initialization is recognized as excellent for deep learning and, in our case, provided excellent results (other kinds of initializations were tested, such as Sigmoid Uniform, Uniform, Relu, and Relu Uniform). As explained by Glorot and Bengio (2010), Xavier

initialization sets a layer's weights to variates from a random uniform distribution that's "bounded between:

$$\pm \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}$$

– where n_i is the number of incoming network connections, or fan-in, to the layer, and n_{i+1} is the number of outgoing network connections from that layer, also known as the fan-out.

- **Second Dense Layer Features**: The second dense layer uses ten neurons, a learning rate of 0.1, an activation function of the sigmoid, and a Xavier type of initialization of weights.
- **The Output Layer**: it has only one output neuron indicating the level of probability and other parameters such as a learning rate of 0.1, a Uniform initialization, and the loss function was set to Sum of Square Errors, which is entirely appropriate for regression problems.

For the optimization algorithm, gradient descent was the best. Gradient descent was combined with a line search, finding the locally optimal step size on every iteration (and we selected a search of 10-line search operations – this resulted in more time for training but assured a better result) (Ibrahim and Rabelo 2021; Glorot and Bengio 2010; Goodfellow et al. 2016). Figure 4 provides details of the best architecture for this experiment.

The next parameters to analyze and try were the batch size, the number of epochs, and the number of iterations. Batch size is the number of samples processed before the model is updated. The number of epochs is the number of complete passes through the training dataset. Finally, iterations are the number of parameter updates done on one input data batch. The best results were given with a minimum of 6 epochs, 10 iterations, and a batch size of 20.

The training was excellent for trials of 5 epochs, and the average prediction error was 0.11% for prediction samples. The results were excellent, and now we can try other things, such as interpolation. The heat maps of Figures 4 and 6 was reproduced, considering X_1, X_2, X_3, X_7, X_8, X_9, X_{10b}, X_{11}, X_{12a}, X_{12b}. Figure 4 and 5 display the heat map for MOP X_1 and X_{12b} (Blue Forces) vs. the entire universe of MOPs of the Red Forces.

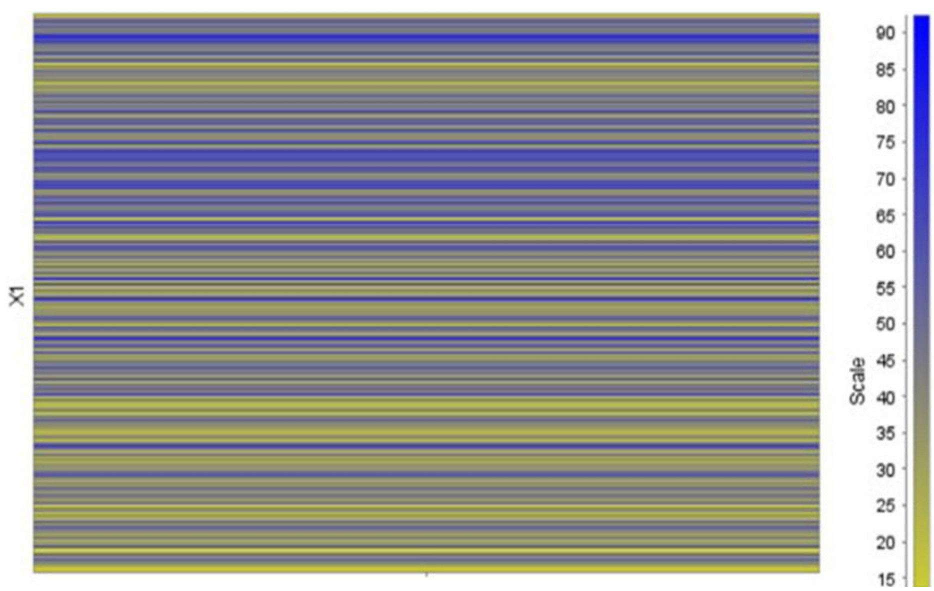

Figure 4: Heat map showing the impact of Blue Team MOP X_1 (number of bombs released) on mission success probability or Measure of Effectiveness (MOE).

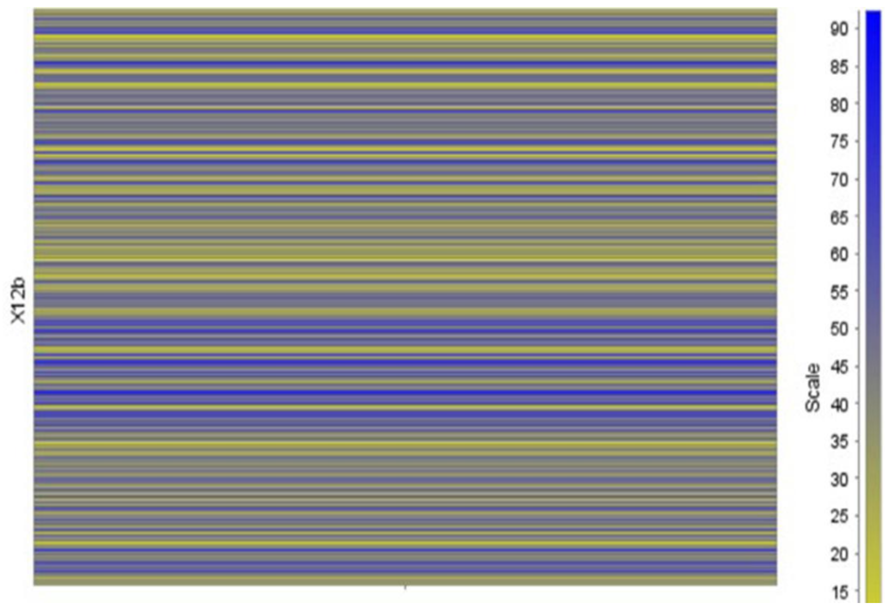

Figure 5: Heat map showing the impact of Red Team MOP X_{12b} (e.g., specific SAM deployment configuration) on Blue Force mission success probability.

5 RESULTS AND VALIDATION WITH SMES

Figures 4 and 5 visually display the MOP versus the probability. The map represented in Figure 4 visualizes the impact of MOP X_1—interpreted as the *number of bombs released by the Blue Force*, an outcome metric relevant to Weapon Combat Effectiveness (WCE), such as the probability of mission success or a composite Measure of Effectiveness (MOE). The Y-axis corresponds to discrete or sampled values of X_1 across hundreds (or thousands) of simulated scenarios. The color gradient, ranging from yellow (low effectiveness, ~15) to blue (high effectiveness, ~90), indicates the computed outcome score based on the deep learning model's evaluation of each configuration.

From a WCE perspective, this heat map is a decision-support visualization tool for mission planners and analysts. The layered horizontal bands represent how effectiveness outcomes vary with changes in the quantity of ordinance released, allowing users to:

- Identify threshold effects: e.g., certain values of X_1 might suddenly yield higher mission success rates (transition from yellow to blue).
- Evaluate diminishing returns: e.g., effectiveness plateaus despite increased bomb deployment, which could signal inefficiency or unnecessary resource expenditure.
- Assess robustness: The alternating bands may indicate that mission success depends not solely on X_1 but on its interaction with other variables (e.g., tactics, Red Force deployment), reinforcing the importance of multi-MOP optimization.
- Support decision trade-offs: By analyzing which values of X_1 result in acceptable effectiveness without overcommitting resources, commanders can balance tactical aggressiveness with risk and resource constraints.

Visual regularity also highlights that X_1 alone is not a sufficient determinant of success; rather, it must be evaluated with other MOPs. As such, this heat map is a diagnostic tool in the broader WCEE development process, offering an empirical foundation for refining tactics, force configuration, and loadout planning under varied operational conditions.

On the other hand, the influence of MOP X_{12b}—associated with a specific Red Team deployment configuration—on a WCE-related outcome, such as the probability of Blue Force mission success or a modeled Measure of Effectiveness (MOE – See Figure 5). The Y-axis represents the full range of simulation instances or configurations involving different values of X_{12b}, while the color gradient (from yellow to dark blue) indicates the output metric on a scale of 15 (low effectiveness) to 90 (high effectiveness):

- X_{12b} likely corresponds to a binary-coded or categorical decision variable indicating whether a particular Red Team SAM or defensive strategy was deployed. Even if it is a binary variable (e.g., 0 or 1), it may have been simulated across many different contexts, reflected in the many lines on the Y-axis.
- Color Bands and Effectiveness Zones:
    - The dominance of blue hues near the middle of the map suggests that scenarios where X_{12b} is active (or in a specific configuration) correlate with lower Blue Team effectiveness, that is, the Red Team defense under this deployment significantly disrupts the mission success.
    - Conversely, lighter zones (yellow-to-green) toward the upper and lower sections suggest more favorable WCE outcomes when X_{12b} is inactive or configured differently.

This heat map serves as a critical situational awareness tool in mission planning and weapon system configuration:

- Red Team Defensive Impact Assessment: The visualization enables planners to quantify how specific Red Team deployments (e.g., $X_{12b} = 1$) diminish mission effectiveness, providing essential input into countermeasure planning, flight path adaptation, or decoy deployment.
- Threat Identification & Avoidance: Lower MOE (darker blue) areas highlight high-threat environments that should be avoided or neutralized early in the mission timeline. This supports prioritizing the targeting of Red SAM systems during early strike phases.
- Tactical Adaptation: The Blue Team can use this intelligence to reconfigure attack packages (e.g., increase chaff/flares, change formation, or alter entry angle) specifically when X_{12b}-type defenses are detected or anticipated.

When viewed alongside the heat map for X_1 (number of bombs released), Figure 5 for X_{12b} provides cross-dimensional insight:

- X_1's map shows the offensive control variable from the Blue Force side.
- X_{12b} reflects defensive influence from the Red Force.

To validate the model's outputs, we consulted Subject Matter Experts (SMEs) and assessed the model's extrapolation capabilities by adjusting the ranges of X_1 (number of bombs released, 1–6) and X_7 (number of missiles, 1–7). As illustrated in Figure 6, variations in X_{11} (number of attack flights) led to anticipated shifts in the probability distribution, with scenarios involving fewer flights showing higher mission effectiveness under certain configurations. The appearance of a yellow band near the top of the map highlights a decline in effectiveness when X_1 exceeds its operational limit (capped at 5). Despite this infeasibility, the deep learning model successfully extrapolated outcomes while maintaining valid probability bounds between 0 and 1, demonstrating robustness and reliable behavior in edge cases.

Upon reviewing these probabilities with our SMEs, they agreed with the results. They explained that there are more opportunities with other types of military problems, such as package selection. They confirmed that it was essential to use unstructured and structured data together. Dealing with noise, nonlinearities, and extrapolation capabilities is essential in this WCE domain.

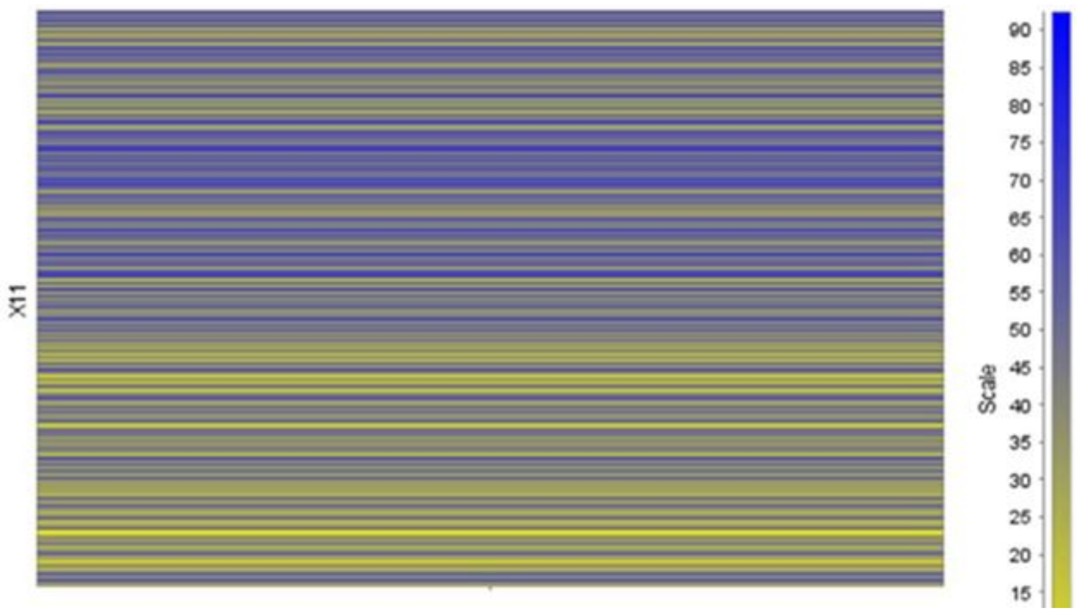

Figure 6: The distribution of probabilities according to the variations of input X_{11} going from small to large values.

6 CONCLUSIONS

This paper revisits the Weapon Combat Effectiveness (WCE) model by Jung (2018), addressing its limitations through advanced data-driven techniques. We went beyond the original regression model and raw data by training and testing deep learning architectures capable of capturing nonlinear dynamics and uncovering hidden patterns in high-dimensional simulation data. The resulting neural networks successfully predicted Measures of Effectiveness (MOEs) and reproduced—while enhancing—the original probability maps.

A key outcome was demonstrating how deep learning overcomes the constraints of linear models, particularly in handling noise, unstructured data, and nonlinear relationships. This allowed the inclusion of previously excluded variables like the number of defensive SAMs (X_{13}), improving prediction accuracy and interpretability.

Future work will focus on integrating package selection optimization**,** where weapons, missiles, and bomb loadouts are dynamically configured to improve mission success, building on Multi-Model Predictive Control (M2PC) principles (Rabelo et al. 2000). Adopting transformers and large language models (LLMs) presents new opportunities to fuse multimodal inputs (e.g., cockpit images, mission logs, and debrief texts), enhancing WCE analytics through richer contextual understanding and improved adaptability.

REFERENCES

Cortes, E., L. Rabelo, A. T. Sarmiento, and E. Gutierrez. 2020. "Design of Distributed Discrete-Event Simulation Systems Using Deep Belief Networks". *Information* 11(10):467.

Epstein, S. L. 1994. "Toward an Ideal Trainer". *Machine Learning* 15:251–277.

Glorot, X. and Y. Bengio. 2010. "Understanding the Difficulty of Training Deep Feedforward Neural Networks". In Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), *JMLR: W&CP* 9, 249–256. Chia Laguna Resort, Sardinia, Italy.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep Learning*. Cambridge, MA: The MIT Press.

Ibrahim, B. and L. Rabelo. 2021. "A Deep Learning Approach for Peak Load Forecasting: A Case Study on Panama". *Energies* 14(11):3039.

Jung, Won. 2018. "A Methodology on Weapon Combat Effectiveness Analytics Using Big Data and Live, Virtual, Or/and Constructive Simulations". Ph.D. thesis, UCF Electronic Theses and Dissertations, Orlando, Florida. https://stars.library.ucf.edu/etd/5889, accessed August 26, 2024.

Jung, W. I., L. Lowe, L. Rabelo, G. Lee, and O. Kwon. 2017. "A Methodology on Guiding Effectiveness-Focused Training of the Weapon Operator Using Big Data and VC Simulations". *SAE International Journal of Aerospace* 10:57–67.

Jung, W. I., M. Marin, K. Lee, L. Rabelo, G. Lee, and D. Noh. 2019. "Weapon Combat Effectiveness Analytics Using Big Data and Simulations: A Literature Review". *SAE International Journal of Advances and Current Practices in Mobility* 1:357–374.

Kim, K., T. Park, J. Pastrana, M. Marin, E. A. Cortes, L. C. Rabelo, and G. Lee. 2014. "Modeling of Complex Scenarios Using LVC Simulation". In *2014 Winter Simulation Conference (WSC)*, 2931–2941 https://doi.org/10.1109/WSC.2014.7020133.

Kim, K., T. W. Park, R. Riad, S. Bhide, O. Joledo, J. Pastrana, M. Marin, L. Lowe, L. Rabelo, and G. Lee. 2014. "Test-Beds Using the High-Level Architecture and Other Distributed-Simulation Frameworks". In *IISE Annual Conference Proceedings*, 1291. Institute of Industrial and Systems Engineers (IISE).

Kirby, B., G. Fletcher, and H. Dudfield. 2011. "Live Virtual Constructive Training Blend Optimisation Study". In *NATO Modelling and Simulation Group Symposium*, October 24-25, 2011, Berne, Switzerland.

Rabelo, L., D. Godbole, J. Jelinek, and V. Gopal. 2000. "Multi-Model Predictive Control: From Air Operations to Enterprise Optimization". In *Proceedings of the DARPA-JFACC Symposium on Advances in Enterprise Control*, July 2000, Minneapolis, Minnesota, 149–157.

Lee, G., J. Kim, M. Marin, K. Lee, E. Gutierrez, and L. Rabelo. 2019. "Building Multiple Resolution Modeling Systems Using the High-Level Architecture". *SAE International Journal of Advances and Current Practices in Mobility* 2:838–842.

Lowe, L., L. Rabelo, M. Elkamel, M. Hunsucker, K. Arias-Marin, N. Davila, O. Allaz, M. Marin, and G. Lee. 2024. "Artificial Intelligence and Simulation for Enhanced Pilot Training". In *2024 Winter Simulation Conference (WSC)*, 1–12 https://doi.org/10.1109/WSC63780.2024.10838989.

Rabelo, L., S. Bhide, and E. Gutierrez. 2018. *Artificial Intelligence: Advances in Research and Applications*. New York: Nova Science Publishers, Inc.

Rabelo, L., E. Gutierrez-Franco, A. Sarmiento, and C. Mejía-Argueta. 2021. *Engineering Analytics: Advances in Research and Applications*. CRC Press.

Rabelo, L., L. Low, W. Jung, M. Elkamel, and G. Lee. 2024. "Enhancing Effectiveness Through the Fusion of Machine Learning, VC Simulation, and Agent-Based Modeling". In *IISE Annual Conference. Proceedings*, 1–6. Institute of Industrial and Systems Engineers (IISE).

Summers, J. E. 2012. "Simulation-Based Military Training: An Engineering Approach to Better Addressing Competing Environmental, Fiscal, and Security Concerns". *Journal of the Washington Academy of Sciences* 9-29.

Tesauro, G. 1995. "Temporal Difference Learning and TD-Gammon". *Communications of the ACM* 38(3):58–68.

West, E. A., C. A. Paul, D. Webb, and W. H. Potter. 2013. "Variation of Instructor-Student Interactions in an Introductory Interactive Physics Course". *Physical Review Special Topics- Physics Education Research* 9(1):010109.

AUTHOR BIOGRAPHIES

LUIS RABELO is a Professor in the Department of Industrial Engineering and Management Systems at the University of Central Florida and Co-Director of the Simulation Interoperability Lab. He is an expert in management systems, artificial intelligence (AI), and simulation. His email address is luis.rabelo@ucf.edu, and his website is https://iems.ucf.edu/faculty/luis-rabelo/.

LARRY LOWE has a Ph.D. in Industrial Engineering from the University of Central Florida. He has worked extensively in Live Virtual Constructive Simulations. His email address is larry.lowe@gmail.com.

WON JUNG received his Ph.D. in Industrial Engineering from the University of Central Florida, where he specialized in VC Simulation and Big Data. His research interests include Combat Modeling and Simulation, Big Data Analytics, and Deep Learning.

MARWEN ELKAMEL is an Associate Researcher in the Department of Industrial Engineering and Management Systems at the University of Central Florida. His research focuses on the Food-Energy-Water (FEW) Nexus, and Energy systems, using Agent-based modeling, Machine Learning, and Operations Research. His email address is marwen.elkamel@ucf.edu.

GENE LEE is an Associate Professor in the Department of Industrial Engineering and Management Systems at the University of Central Florida. He is the Director of the Simulation Interoperability Laboratory. His email is glee@ucf.edu.