Comparing Backpropagation-based Neural Networks and Deep Learning to Predict the NASA Task Load Index

Gabriel Perez and Juanita Gonzalez
Department of Industrial Engineering & Management Systems
University of Central Florida
gabrielpz@knights.ucf.edu, juanitag@knights.ucf.edu

Dr. Luis Carlos Rabelo
Department of Industrial Engineering & Management Systems
University of Central Florida
luis.rabelo@ucf.edu

Abstract

Virtual Reality and Deep Learning are technologies that can enhance Human Engineering. This paper discusses some experiments using paradigms of neural networks (Machine Learning) vs. Deep Learning to predict the NASA Task Load Index (TLX) from general information and brain waves without answering the questionnaire. The environment of the experiments was a virtual reality-based one using OpenSim to build it and Oculus Rift to navigate it. Electroencephalograms were used to measure the level of engagement of the users. After the tasks were performed, the TLX questionnaire was completed. Resilient Backpropagation and Deep Learning-based architectures were used to create a mapping of brain waves and demographic information to the TLX index. The results are very positive, and the TLX index can be predicted. These results can support the development of real-time assessment systems and the build-up of adaptable/smart user interfaces.

Keywords
TLX index, Deep Learning, RProp, Neurons, Mental Workload Assessment

1.Introduction
The Human engineering field supports the appropriate evolutions of technology to adapt it to human use. The main focus is on the adaptability of hardware, software, and systems. Providing humans with the most effective tools will allow them to improve their capability and performance when performing a task. At the same time, it can help ensure a better level of safety. In this case, this discipline is a key factor of this research since it involves studying human factors related to the workload assessment. The mental workload assessment based on the performance of a task is measured based on the National Aeronautics and Space Administration Task Load Index (NASA TLX) questionnaire (Rubio et al., 2004). The rating is based on six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration (Grier et al., 2003).

By predicting the mental workload, engineers will be able to find different strategies and the best ways to focus mental workload on fundamental tasks. Doing so will increase, at the same time, the level of attention that is given to a certain task. TLX allows one to identify key factors that can influence a person's performance to change it and accommodate the necessary adjustments. An increase in productivity and effectiveness will result from such changes in better allocation of tasks.

2. AI, Machine Learning, Deep Learning
Artificial Intelligence (AI) has many definitions that have increased over more than 6 decades. However, Dr. John McCarthy first used the term AI (Researcher at MIT and Professor at Stanford University) in 1955. McCarthy defined AI as that science and engineering field that designs and makes intelligent machines (Sutton, 2020). So, one of the most important features is that machines can learn in the same way (or better) as humans.

One of the most important branches of AI is machine learning (ML). ML emphasizes computer agents can improve their knowledge or actions based on experience. This experience can mean a lot of data. ML is an interdisciplinary field based on computer science, statistics, psychology, electrical engineering, industrial engineering, neurosciences, and mathematics (Rabelo et al., 2018). The data is very important so that the machine/algorithm learns by itself from the data. Neural networks are part of ML.

One of the relatively recent advances in ML is Deep Learning. Deep Learning is a neural network that uses many layers that compute with representations suitable for its hierarchical structure resembling the human brain. Deep Learning has been very successful, especially with pattern recognition, images, and speech processing problems. Furthermore, Deep Learning works well with limited data and very much data. In particular, it is very adaptable to problems with Big Data, where the structured is mixed with the unstructured.

It is appropriate to say that several of these schemes must always be compared to choose the one that best suits our problem. And from there is the reason for this article where we make comparisons of ML as traditional neural networks like Resilient Backpropagation (RPROP) against Deep Learning and thus choose the most appropriate for our problem: Map demographic data and brain waves to TLX.

Resilient Backpropagation (RPROP)
A variant of the Backpropagation Neural Networks based on gradient descent is Resilient backpropagation or RPROP. It is an algorithm used to train neural networks and is similar to the commonly used backpropagation method. However, it has two advantages, it is faster and does not require specifying parameter values (Riedmiller & Braun, 1993).

RPROP has been widely used since it does not require an initial learning rate like the backpropagation model. The main difference with the backpropagation model is that the partial derivatives of the error function are used only to correct the direction of the network weights. This utilization of partial derivatives makes it faster than traditional backpropagation models.

The RPROP technique has proven to tolerate training databases containing noise. For this reason, its use has been extended to the analysis of situations with empirical data. RPROP is adequate for the task of prediction of the TLX index.

Deep Learning (DL)
Deep Learning is the latest development of ML (Patterson & Gibson, 2017). The universe of Deep Learning has many different architectures and algorithms (Cortes et al., 2020; Ibrahim & Rabelo, 2021). Cortes et al. (2020) stated that: "Deep learning is self-learning by constructing a model with several layers and training it with data. This nature of multiple layers can improve the accuracy of the classification." The one utilized in this research has many layers and utilizes different optimization schemes such as ADAM and the classical Stochastic Gradient Descent. Moreover, dense layers are utilized to take advantage of the full connectivity.

3.Virtual Reality and Brain's Electrical Signals Environment/Experiments

The creation of a 3D model in immersive virtual reality (IVR) for simulation training purposes has the potential of having the trainees live an experience that is closer to a real-life scenario. Therefore, the virtual environment created for this research sought to model the Engineering Building II at UCF, where the simulation training experiments took place. During these experiments, the team involved in designing the virtual environment decided to use OpenSimulator (Bhide 2017) as their designated software tool due to its quality, cost, and previous uses in the academic field, where it has delivered excellent results. The most significant reason that prompted the use of OpenSimulator was the persistence, stability, and portability of the virtual environment simulation. It allows the simulation to be persistent since it keeps the state of the scenario or data even after the device is turned off. In addition, it is an open-source platform that allows a secured exchange of data and communication with external systems (Bhide, 2017).

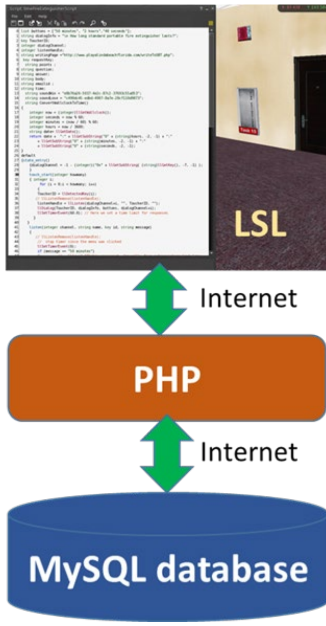

Figure 1. DataBase and OpenSim relation (Adapted from Bhide 2017 with permission)

The creation, manipulation, and behaviors of the primitive objects used in the simulated environment are developed using scripts of the programming language called Linden Scripting Language (see Figure 1). For example, if the creator wants to upload their content to the platform, it can be easily done. In addition, MySQL provided data storage functionality for future analysis and training feedback. Also, PHP was used for connection purposes between both platforms.

The experiments were performed using a VR-ready Laptop. The EMOTIV EPOC hardware was used to document the brain's electrical signals from the trainees (see Figure 2). Electrodes were placed on the head of the participants without interrupting the person's movement being evaluated. The green signals from the Testbench software indicated the assertive contact placement of the electrodes to detain the required information.

One of the most important apparatuses used during this experiment was the Virtual Reality Oculus Rift set. It allowed the user to have the "real-world experience" of the simulated environment. Along with the goggles, the electroencephalogram (EEG) headset is placed on the participants to record their engagement.

Figure 2. EMOTIV EPOC+, Oculus Rift, and VR-ready laptop (Adapted with permission from Alasim 2020)

4. Environment Scenarios & Tasks

Two virtual scenarios were designed with 5 main differences to collect the data for the TLX experiment. Hence, it is important to mention that these changes were done to evaluate the factors that affect attention in human beings based on cognitive science (Alasim, 2017). In addition, the virtual environments created for the experiment added several important characteristics of the building, such as windows, stairs, floors, walls, ceiling, and textures of objects and carpets to produce conditions that were as close as possible to the real premises. Furthermore, the location of halls and stairs, which play a critical role in emergency cases, were also designed just like the ones in the building (see Figure 3).

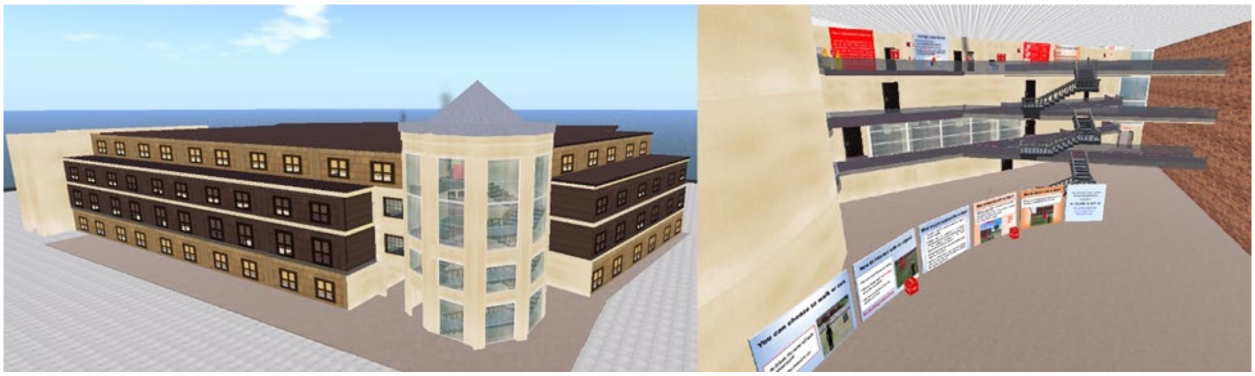

Figure 3. Virtual Engineering Building and Atrium

When the trainees were ready to enter the virtual world, their first step was to select an avatar. During this process, the users were given two choices of avatars: a male or a female. In addition, it is important to mention that the researchers previously designed these two avatars before the participants entered the experiment. Furthermore, after selecting their desired avatar, the participants could observe the entire virtual environment using the Oculus Rift goggles and navigate it as desired. Also, the avatars in the virtual environment can teletransport, walk, run, fly, and even interact with objects by using the computer's keyboard and mouse.

Once the participants are ready to begin their training, they walk to the building's atrium, where they are presented with an instruction board for the tasks they will perform during the experiment. Consequently, participants must read and understand the instructions in this initial stage before proceeding to the next part of the experiment. Finally, after everything is clear, the participants can click on the posters and be teletransported to the 4th floor, where all the tasks for the experiment take place.

After being teletransported to the fourth floor, the participants will perform 15 different tasks. Both environments have the same amount of tasks but with changes in the format for instructions and responses. Here, the tasks are presented on the walls, doors, and corridors of the fourth floor, and participants must make decisions and interact with the objects in each of the 15 tasks. Finally, the last task informs the participant that the training has been completed and that they can proceed to the exit. If the participants are in environment 2, they will hear an alarm until they find the only exit available. The diagram below represents the changes done in the environments.

4.1Experimentation with Neural Networks and Deep Learning

After collecting data on the brain's electrical signals from the participants in the VR training experiment, we began assembling our neural networks using Knime (www.knime.com). Knime is an open-source data analytics platform that stands for Konstanz Information Miner, and it is the tool we used to build our models and train neural networks. Furthermore, this software provided us with tables and graphs for our analysis (see Figure 4).

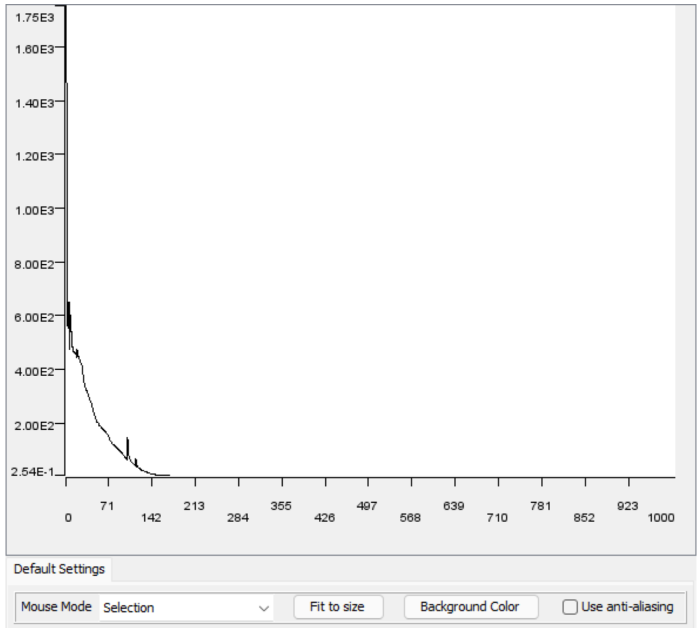

Figure 4. Error Plot for RPROP Training

After downloading the program and its extensions for deep learning, we could assemble our models to start training our gathered data. First, we trained the RPROP models using 1, 2, and 3 layers and tested each layer with 5, 10, and 20 neurons except for layer 3, where we also tested for 25 and 30 neurons due to finding positive results at 20 neurons. Second, we trained the Deep Learning models using 2, 3, and 4 layers and tested each one with 20, 50, 100, and 200 neurons. In total, we built 48 models for Deep Learning, 11 for RProp, and 2 models for the final analysis.

5. Result and Discussion

After doing the experiments, we trained the neural networks with different configurations looking for the neural network with the lowest error. Hence, we shortened the actual data of 11822 rows to 10000 for training only and left the rest of the data for validation. Furthermore, we used the formulas for Mean Absolute Error (MAE) and the Mean Squared Error (MSE) to check the accuracy of the predictions in the training process and the validation. In addition, we decided to use MAE to calculate how much error could be identified in the prediction and MSE to identify the loss of the squared error and determine the proximity of the regression line to the scores. Finally, those results helped us determine that the experiment was done with three layers and 20 neurons; both Deep Learning (DL) and RPROP had the least percentage error, categorizing them as the winner.

Table 1. MAE and MSE validation results for each configuration (unseen data)

Configuration Selected Using RPROP & Validation File		
CONFIGURATION	**MAE**	**MSE**
ONE LAYER		
5 Neurons	0.0407870	0.0031690
10 Neurons	0.0176760	0.0007070
20 Neurons	0.0069650	0.0001580
TWO LAYERS		
5 Neurons	0.0406450	0.0043300
10 Neurons	0.0072020	0.0001550
20 Neurons	0.0039900	0.0000700
THREE LAYERS		
5 Neurons	0.0363370	0.0070970
10 Neurons	0.0061330	0.0001350
20 Neurons	0.0035080	0.0000310
25 Neurons	0.0037320	0.0000340
30 Neurons	0.0043240	0.0000510

Table 2. MAE and MSE results from configuration using Deep Learning

Configuration Using Deep Learning		
CONFIGURATION	**MAE**	**MSE**
TWO LAYERS		
20 Neurons	0.1666000	0.1144151
50 Neurons	0.2939000	5.4263547
100 Neurons	0.1891000	0.1657011
200 Neurons	0.4317	0.4720064
THREE LAYERS		
20 Neurons	0.0799000	0.0379407
50 Neurons	0.1651000	0.0932780
100 Neurons	0.0838000	0.0391555
200 Neurons	0.1755100	0.1532266
FOUR LAYERS		
20 Neurons	0.0961000	0.0398236
50 Neurons	0.2329000	0.1617614
200 Neurons	0.1876000	0.2080361
25 Neurons	0.2881000	0.3388172

After finding the winners for RProp and DL, we moved forward to the validation process, where we used the trained neural networks to predict the data of the 11822 rows. We also used the MAE and MSE as references for finding the lowest percentage of error during this process. This process helped us decide which neural network gives the best results in predicting the NASA TLX.

To complete our experiment, we configured the predictors of both neural networks to receive the already trained neurons and the validation data separately as inputs. Therefore, the module could produce a new prediction for the 11822 data and complete the validation process. Our final experiment with the whole data led us to believe that the Resilient Backpropagation (RProp) was better at predicting the NASA TLX. The following chart shows the MAE and MSE results found in both RProp and DL:

Table 3. MAE and MSE results found in RProp and Deep Learning

	RProp	**Deep Learning**
MAE	0.02219196443	0.9632841036
MSE	8.85E-04	0.2214592456

In the future, including in this research, the Presence Questionnaire (PQ) finding will offer a variety of factors to contribute to the study of mental workload. First, it contributes to evaluating the experience of being in a Virtual Environment for a person. This evaluation will help find pertinent factors that affect a person's attention depending on their surrounding conditions.

Acknowledgments

A scholarship supported Juanita. This research was the result of Undergraduate Research EIN4912 for both authors. Furthermore, the authors would like to thank Dr. Luis Rabelo for the invitation to participate in his group as part of the UCF IEMS Simulation Interoperability Laboratory. Furthermore, the authors would like to thank Dr. Sayli Bhide (Indian Institute of Technology (IIT) Bombay, Mumbai, India) and a member of the research group for the teachings in Virtual Reality and Dr. Fahad (King Saud University, Riyadh, Saudi Arabia) for his teachings and using his data for the TLX.

References

Alasim, F. , Human Performance In Virtual Reality Environments and Its Exploration with Engineering Analytics (Doctor of Philosophy). The University of Central Florida).2020

Bhide, S, Fire Safety and Emergency Evacuation Training for Occupants of Building Using 3D Virtual Simulation. (Doctor of Philosophy). The University of Central Florida.2017

Cortes, E., Rabelo, L., Sarmiento, A. T., & Gutierrez, E, Design of Distributed Discrete-Event Simulation Systems Using Deep Belief Networks. Information, 11(10), 467.2020.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R, The vigilance decrement reflects limitations in effortful attention, not mindlessness. Human factors, 45(3), 349-359.2003.

Ibrahim, B., & Rabelo, L, A deep learning approach for peak load forecasting: A case study on Panama. Energies, 14(11), 3039.2021.

Patterson, J., & Gibson, A, Deep learning: A practitioner's approach: " O'Reilly Media, Inc.".2017.

Rabelo, L., Bhide, S., & Gutierrez, E. , Artificial intelligence: Advances in research and applications: Nova Science Publishers, Inc.2018

Riedmiller, M. & Braun, H, A direct adaptive method for faster backpropagation learning: The Rprop algorithm. Proceedings of the IEEE International Conference on Neural Networks, 586-591, IEEE Press, 1993

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. , Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. Applied Psychology, 53(1), 61-86, 2004.

Sutton, R. S, John McCarthy's definition of intelligence. Journal of Artificial General Intelligence, 11(2), 66-67.2020.

Biographies

Gabriel Perez current Industrial Engineering Student at the University of Central Florida graduating in December. He has worked as an Engineer Intern for Café del Sur and now as an Area Manager Intern for Amazon. His research interests include supply chain, neural networks, and artificial intelligence.

Juanita Gonzalez is an Industrial Engineering student at the University of Central Florida, graduating in 2022. Supply Chain Intern at PepsiCo. Research interests include Deep Learning and Neural Networks.

Dr. Luis Rabelo is a Professor in the Department of Industrial Engineering and Management Systems at the University of Central Florida.