



**2023 MSVSCC
Paper Abstracts**



Data Science Paper Abstracts

Assessing the Frequency and Severity of Malware Attacks: An Exploratory Analysis of the Advisen Cyber Loss Dataset

Ahmed Abdelmagid, Farshid Javadnejad, C. Ariel Pinto, Michael McShane, Rafael Diaz, Elijah Gartell

In today's business landscape, cyberattacks present a significant threat that can lead to severe financial losses and damage to a company's reputation. To mitigate this risk, it is essential for stakeholders to have an understanding of the latest types and patterns of cyberattacks. The primary objective of this research is to provide this knowledge by utilizing the Advisen cyber loss dataset, which comprises over 137,000 cyber incidents that occurred across various industry sectors from 2013 to 2020. By using text mining techniques, this paper will conduct an exploratory data analysis to identify the most common types of malwares, including ransomware. Furthermore, the study will include a likelihood and severity analysis to evaluate the financial impact of these cyberattacks on businesses. Ultimately, this study aims to shed light on the prevalence and financial repercussions of malware incidents and provide businesses with valuable insights to help develop effective cybersecurity strategies.

Keywords: Ransomware, Malware, Cyber-attack, Advisen data loss, Likelihood, Severity

The Effectiveness of Visualization Techniques for Supporting Decision-Making

Cansu Yalim, Holly Handley

Although visualization is beneficial for evaluating and communicating data, the efficiency of various visualization approaches for different data types is not always evident. This research aims to address this issue by investigating the usefulness of several visualization techniques for various data kinds, including continuous, categorical, and time-series data. The qualitative appraisal of each technique's strengths, weaknesses, and interpretation of the dataset is investigated. The research questions include: which visualization approaches perform best for different data types, and what factors impact their usefulness? The absence of clear direction for both researchers and practitioners on how to identify the most effective visualization approach for a specific data type poses a significant research challenge. Our findings will help both professionals and researchers determine the most successful visualization approach for different data types, as well as identify topics for future study in the field of data visualization.

Keywords: Data analysis, Decision-making, Information visualization, Visualization Techniques, Data communication, Data interpretation

Data Science Paper Abstracts

Extracting Information from Twitter Screenshots

Tarannum Zaki

Screenshots are prevalent on social media as a common approach for information sharing. Users rarely verify before sharing screenshots whether they are fake or real. Information sharing through fake screenshots can be highly responsible for misinformation and disinformation spread on social media. There are services of the live web and web archives that could be used to validate the content of a screenshot. We are going to develop a tool that would automatically provide a probability whether a screenshot is fake by using the services of the live web and web archives.

Keywords: Twitter, misinformation, disinformation, screenshot, web archives.

Behind Derogatory Migrants' Terms for Venezuelan Migrants: Xenophobia and Sexism Identification with Twitter Data and NLP

Joseph Martinez, Melissa Miller-Felton, Jose Padilla, Erika Frydenlund

The sudden arrival of many migrants can present new challenges for host communities and create negative attitudes that reflect that tension. In the case of Colombia, with the influx of over 2.5 million Venezuelan migrants, such tensions arose. Our research objective is to investigate how those sentiments arise in social media. We focused on monitoring derogatory terms for Venezuelans, specifically *veneco* and *veneca*. Using a dataset of 5.7 million tweets from Colombian users between 2015 and 2021, we determined the proportion of tweets containing those terms. We observed a high prevalence of xenophobic and defamatory language correlated with the terms, with the latter perpetuating a sexist portrayal of female Venezuelan migrants.

Keywords: Migration, xenophobia, natural language processing, social media.

The Legacy of Colonization and Civil Societies in South Africa

Erika Frydenlund, Melissa Miller-Felton, Bolu Ayankojo

This research analyzes the unique ways that civil societies operate in Sub-Saharan Africa in the context of post-apartheid Cape Town, South Africa. Decades after the demise of apartheid, remnants of inequality remain without the promise of actionable change. We used a computational modeling approach to understand the dynamics of migrants in the receiving community as derived from qualitative interviews conducted with 24 stakeholders in Cape Town, South Africa between 2020 and 2021. Our findings show that the presence of NGOs can promote access to resources and reduce xenophobia if they can have the right influence on government policies.

Keywords: NGOs, Khayelitsha, Migrants, Xenophobia, Systems Modeling

Data Science Paper Abstracts

Assessing Frustration Towards Venezuelan Migrants in Colombia: Path Analysis on Newspaper Coded Data

Brian Llinas, Guljannat Huseynli, Erika Frydenlund, Katherine Palacio,
Jose Padilla

This study analyzes the impact of Venezuelan migrants on local frustration levels in Colombia. The study found a relationship between the influx of Venezuelan migrants and the level of frustration among locals towards migrants, infrastructure, government, and geopolitics. Additionally, we identified that frustration types have an impact on other frustrations. The study used articles from a national newspaper in Colombia from 2015 to 2020. News articles were coded during a previous study qualitatively and categorized into frustration types. The code frequencies were then used as variables in this study. We used path modeling to statistically study the relationship between dependent and independent variables through mediator variables. This paper aims to fill the research gap by contributing a unique model to the literature and insights about host community attitudes towards migrants and national migration response.

Keywords: Mixed Methods, Frustration, Migrants, Path Model, News Analysis

Exploring Xenophobic Events through Gdelt Data Analysis

Hirmarsha Jayanetti, Erika Frydenlund, Michele Weigle

This study explores xenophobic events related to refugees and migration using the GDELT 2.0 database and APIs through visualizations. We conducted two case studies – the first being an analysis of refugee related news following the death of a two-year-old Syrian boy, Alan Kurdi, and the second a surge in news articles in March 2021 based on the data obtained from GDELT API. In addition to the two case studies, we present a discussion of our exploratory data analysis steps and the challenges encountered while working with GDELT data and its tools.

Keywords: Xenophobia, Refugees and Migrants, GDELT, Big Data, Data Science for Social Good

GPU Utilization: Predictive SARIMAX Time Series Analysis

Dorothy Parry

This work explores collecting performance metrics and leveraging the output for prediction on a memory intensive parallel image classification algorithm - Inception v3 (or "Inception3"). Experimental results were collected by nvidia-smi on a computational node DGX-1, equipped with eight Tesla V100 Graphic Processing Units (GPUs). Time series analysis was performed on the GPU utilization data taken, for multiple runs, of Inception3's image classification algorithm (see Figure 1). The time series model applied was Seasonal Autoregressive Integrated Moving Average Exogenous (SARIMAX).

Keywords: Time Series, Machine Learning, GPU Utilization, Performance Measures

General Sciences & Engineering

Statistical Approach to Quantifying Interceptability of Interaction Scenarios for Testing Autonomous Surface Vessels

Benjamin Hargis, Yiannis Papelis

This paper presents a probabilistic approach to quantifying interceptability of an interaction scenario designed to test collision avoidance of autonomous navigation algorithms. Interceptability is one of many measures to determine the complexity or difficulty of an interaction scenario. This approach uses a combined probability model of capability and intent to create a predicted position probability map for the system under test. Then, intercept-ability is quantified by determining the overlap between the system under test probability map and the intruder's capability model. The approach is general; however, a demonstration is provided using kinematic capability models and an odometry-based intent model.

Keywords: Evaluation Metrics, Avoid-ability, Simulation

An Algorithm for Finding Data Dependencies in an Event Graph

Erik Jensen

This work presents an algorithm for finding data-independent events in a discrete-event simulation, from an event-graph model of the system. The algorithm can be used within a parallel discrete-event simulation. Also presented is an experimental system and event graph, which is used for testing the algorithm. Results indicate that the algorithm can provide information about which vertices in the experimental event graph can affect other vertices, and the minimum amount of time in which this interference can occur.

Keywords: Parallel Discrete-Event Simulation, Event Graphs, Data Dependencies, Synchronization Algorithm

Origami-Inspired Declarative Surface Folding Model

Nathaniel Dearce, Andrey Chernikov

Origami has many applications where a large surface area must be compacted to minimize volume. The proposed approach to model surface folding is based in lightweight formal methods. Alloy's specification language is used to model folds of a surface consisting of faces to show there is a satisfiable sequence of folds resulting in a desired configuration.

Keywords: Lightweight Formal Methods, Origami, Rigid Origami

General Sciences & Engineering

U-Net Based Multiclass Semantic Segmentation for Natural Disaster Based Satellite Imagery

Nishat Ara Nipa

Satellite image analysis of natural disasters is critical for effective emergency response, relief planning, and disaster preparation. Semantic segmentation is believed to be one of the best techniques to capture pixelwise information in computer vision. In this work we will be using a U-Net architecture to do a three-class semantic segmentation for the Xview2 dataset to capture the level of damage caused by different natural disasters which is beyond the visual scope of human eyes.

Keywords: Natural Disaster, Damage Assessment, Semantic Segmentation, U-Net

Towards NLP-Based Conceptual Modeling Frameworks

David Shuttleworth, Jose Padilla

This paper presents preliminary research using Natural Language Processing (NLP) to support the development of conceptual modeling frameworks. NLP-based frameworks are intended to lower the barrier of entry for non-modelers to develop models and to facilitate communication across disciplines considering simulations in research efforts. NLP drives conceptual modeling in two ways. Firstly, it attempts to automate the generation of conceptual models and simulation specifications, derived from non-modelers' narratives, while standardizing the conceptual modeling process and outcome. Secondly, as the process is automated, it is simpler to replicate and be followed by modelers and non-modelers. This allows for using a common process and generating similar "blueprints" facilitating communication and collaboration efforts. Overall, NLP presents an opportunity for the M&S community to engage with stakeholders and scholars across domains in the simulation development process, lowering entry barriers and increasing participation.

Keywords: Conceptual Modeling, Conceptualization, Natural Language Processing

Review of Nighttime Temperature Effects on Long-Term Health Conditions Through Sleep Studies

Sydney Matkins

Over the past 40 years, there has been increasing interest in human sleep quality and duration. This nonsystematic review looked at over 80 peer-reviewed papers on the association among sleep, temperature, and long-term health conditions. Generally, warmer temperatures lead to poorer sleep quality, and poor sleep quality leads to mental illness and a higher risk of coronary heart disease and mortality. Future research should be to conduct a study that relies more on health records rather than questionnaires to accurately map current and future health quality.

Keywords: Sleep Quality, Sleep Duration, Thermal Environment, Mental Health, Long-Term Health Conditions

Automatic Generation of Virtual Work Guide for Complex Procedures: A Case Study

Shan Liu, Yuzhong Shen

Practical work guides for complex procedures are significant and highly affect the efficiency and accuracy of on-site users. This paper presents a technique to generate virtual work guides automatically for complex procedures. Firstly, the procedure information is extracted from the electronic manual in PDF format. And then, the extracted procedure steps are mapped to the virtual model parts in preparation for animation between adjacent steps. Next, smooth animations of the procedure are generated based on a 3D natural cubic spline curve to improve the spatial ability of the work guide. In addition, each step's annotation is automatically adjusted to improve the visual effect of the work guide. A troubleshooting procedure example of the M16A4 Rifle shows that the generated work guide is well instructive, which provides the interactive simulation of the procedure and process-based display of technical annotation.

Keywords: Work Guide, Virtual Environment, Procedure Animation, Procedure Annotation

Virtual Environments & Visualization

Local Positioning in a Virtual Environment

Collin Moorefield, Grant Meadows, Laurenz Advincula, Daniel Conrad, John Billy

The technology of Augmented and Virtual Reality are developing rapidly, permeating its advantageous features through vast utilization. The design objective of this proposal is to construct a working application for communicating the awareness between a physical and virtual environment synchronously. Utilizing the Pozyx anchors, tags, and positioning system, the object is tracked for its position and orientation in real time. The project framework provides further options for the user and should be applied in software development for engineering design purposes. The basic concepts and features of our program were tested with a simple pong game, providing sufficient outputs that show the accuracy of our readings. We encourage further testing in different areas and size of space, as that is one of the configurations that will change the accuracy. Future users and organizations that intend on using our framework will have an effortless experience in connecting their positioning systems with a single or multiple environment simulation thread. (Received in April 2023, accepted in April 2023. This paper was with the authors 2 months for 2 revisions.)

Keywords: Pozyx, Augmented Reality, Virtual Reality, Framework, Simulation, Modeling

Lidar Buoy Detection for Autonomous Marine Vessel Using Pointnet Classification

Christopher Adolphi, Dorothy Parry, Ahmet Saglam, Yaohang Li, Masha Sosonkina, Yiannis Papelis

Maritime autonomy, specifically the use of autonomous and semi-autonomous maritime vessels, is a key enabling technology supporting a set of diverse and critical research areas, including coastal and environmental resilience, assessment of waterway health, ecosystem/asset monitoring and maritime port security. Critical to the safe, efficient and reliable operation of an autonomous maritime vessel is its ability to perceive on-the-fly the external environment through onboard sensors. In this paper, buoy detection for LiDAR images is explored by using several tools and techniques: machine learning methods, Unity Game Engine (herein referred to as Unity) simulation, and traditional image processing. The Unity Game Engine (herein referred to as Unity) simulation data was used for the training and testing of a Pointnet neural network model while the labeled real-world maritime environment point cloud data was used for the model validation. Fitting the Pointnet model on the simulation data, after some data alignment with the LiDAR images allowed for accurate classification of buoys on the real-world data with the 93% of accuracy. A traditional image processing approach using 2D occupancy maps to detect the buoys by shape was used as well and is outlined in the paper.

Keywords: LiDAR, Point Cloud, Machine Learning, Pointnet

Virtual Environments & Visualization

Enhancing Pedestrian-Autonomous Vehicle Safety in Low Visibility Scenarios: A Comprehensive Simulation Method

Zizheng Yan, Yang Liu, Hong Yang

Self-driving cars raise safety concerns, particularly regarding pedestrian interactions. Current research lacks a systematic understanding of these interactions in diverse scenarios. Autonomous Vehicle (AV) performance can vary due to perception accuracy, algorithm reliability, and environmental dynamics. This study examines AV-pedestrian safety issues, focusing on low visibility conditions, using a co-simulation framework combining virtual reality and an autonomous driving simulator. 40 experiments were conducted, extracting surrogate safety measures (SSMs) from AV and pedestrian trajectories. The results indicate that low visibility can impair AV performance, increasing conflict risks for pedestrians. AV algorithms may require further enhancements and validations for consistent safety performance in low visibility scenarios.

Keywords: Autonomous Vehicles, Pedestrians, Virtual Reality, CARLA Simulator, Conflict Risk, Simulation, Safety

Medical Simulation

Multi-Material, Approach-Guided, Controlled-Resolution Breast Meshing For Fe-Based Interactive Surgery Simulation

Motaz Alqaoud, John Plemmons, Eric Feliberti, Krishnanand Kaipa, Gabor Fichtinger, Yiming Xiao, Tanweer Rashid, Michel Audette

This paper proposes a guided, controlled resolution framework for 3D multi-material meshing. Using data from magnetic resonance (MR) images, we efficiently focused on demonstrating our framework for patient specific breast cases. As a result, we can preserve the shared boundaries and enhance the resolution without negating the aspect of simulation computing time needed for finite element analysis (FEA). Our output is a high-quality volumetric mesh comprising 21K cells representing the three main parts for breast surgery simulation and planning, fat, fibroglandular (FGT), and tumor mass. Our approach combines three steps, surface meshing, surface mesh decimation, and generating a volumetric mesh. We showed experimental results for every stage and compared our final output to other literature, proving our method's efficiency in an accurate, simple, and high-quality presentation of a patient-specific breast meshing.

Keywords: Multi-Material Breast Mesh, Shared Boundaries, Simulation, Breast Surgery, Surgery Planning

A Comparison of Nonverbal and Paraverbal Behaviors in Simulated and Virtual Patient Encounters

Sarah Powers, Mark Scerbo, Matthew Pacailler, Macy Kisiel, Baillie Hurst, Ginger Watson, Lauren Hamel, Fred Kron

The present study assessed whether trainees display similar nonverbal and paraverbal behaviors when interacting with a simulated (SP) and virtual patient (VP). Sixty second slices of time following four interactions were rated for the presence and frequency of three nonverbal and paraverbal behaviors. Results revealed that students exhibited fewer behaviors in the VP interaction, possibly due to differences social inhibition or fidelity between the two formats.

Keywords: Nonverbal, Paraverbal, Simulated Patient (SP), Virtual Patient (VP), Thin Slice

Medical Simulation

Enhancement of Deep Learning Protein Structure Prediction

Ruoming Shen

Protein modeling is a rapidly expanding field with valuable applications in the pharmaceutical industry. Accurate protein structure prediction facilitates drug design, as extensive knowledge about the atomic structure of a given protein enables scientists to target that protein in the human body. However, protein structure identification in certain types of protein images remains challenging, with medium resolution cryogenic electron microscopy (cryo-EM) protein density maps particularly difficult to analyze. Recent advancements in computational methods, namely deep learning, have improved protein modeling. To maximize its accuracy, a deep learning model requires copious amounts of up-to-date training data. This project explores DeepSSETracer, a software tool that uses deep learning to predict protein secondary structures in medium resolution cryo-EM density maps of protein samples. Python scripts were created to automate data acquisition tasks for DeepSSETracer. Furthermore, the Python library PDBx was used to parse mmCIF protein files. mmCIF is a relatively new file type that stores experimentally derived atomic models of proteins, and they have begun to replace the conventional PDB file type as the standard for atomic models. This project culminated in making ChainChopper, a program in DeepSSETracer, compatible with the mmCIF file type.

Keywords: Alpha Helix, Beta Sheet, DeepSSETracer, ChainChopper, Cryo-EM, MmCIF file

Education & Training

Simulating Function Generators and Oscilloscopes in a Virtual Laboratory Environment

Yiyang Li, Yuzhong Shen, Charles I. Sukenik

This paper discusses the development of a virtual laboratory for simulating electronic instruments commonly used in science and engineering courses, such as function generators and digital storage oscilloscopes. Mathematical equations are used to represent continuous signals and ensure signal integrity, while C# delegates are adopted to enable communication between simulated devices. The approach allows for loose coupling between software components and high cohesion of individual components and can be applied to other virtual laboratory developments. The virtual laboratory provides a means for students to gain hands-on experience with electronic instruments and improve their understanding of theoretical concepts.

Keywords: virtual environment, education, simulation

Hampton Roads' Building Resilient Communities Flood Game

Gul Ayaz, Katherine Smith, Rafael Diaz, Josh Behr

As rising sea levels and subsequent recurrent flooding disproportionately affects coastal areas, it is crucial to develop a heightened awareness of the impacts of natural disasters on communities and the environments they live in. The Hampton Roads' Building Resilient Communities (BRC) Flood Game is a simulation role-playing game designed to allow players to increase their understanding of the impact of various community response interventions to sea level rise and recurrent flooding. Players will examine and assess the tradeoffs of resiliency investments, the impact policies may have on the population, and the amount of time return on investment takes. The BRC game was designed using research-based game design principles to positively impact the user experience and facilitate the development of the overall game environment.

Keywords: Natural Disasters, Resiliency, Policy Makers, Game Design

Education & Training

Mask-On Breathing Awareness Trainer (mobat)

Robert Larson, Yuzhong Shen

Currently there is no low-cost solution to teach large quantities of Naval Aviators on how to correctly breath on an oxygen mask. This report will discuss the design and implementation of a mask-on breathing apparatus for use with United States Naval Aircrew for initial and annual familiarization. With the current uptick in physiological episodes within the naval aviation enterprise, there has been a need to find solutions to this problem. Currently when a naval aircrew is trained in aviation physiological threats, they do so with normobaric hypoxia tools in the form of the 9A19 Normobaric Hypoxia Trainer for aircrew who do not normally wear a mask during flight and the 9A17 Mask-on Breathing Trainer which is for aircrew who do wear a mask throughout an entire flight. This training evolution only lasts around 15 minutes for each student and does not go over the proper breathing techniques to use to breath while on mask provided oxygen. It would take too long and would be overly costly to use the existing training methods to perform this task. The goal of this project is to rectify the gap in training by providing a low cost means to give positive feedback training to large numbers of students at once.

Keywords: Hypoxia, Naval Aviation, Aviation Training, Aerospace Physiology

Digital Game Based Learning Approach to Math Learning for Students

Gul Ayaz, Katherine Smith

Mathematics is an important subject that is pervasive across many disciplines. It is also a subject that has proven to be challenging to both teach and learn. Students face many challenges with learning math such as a lack of motivation and anxiety. To address these challenges, game-based learning has become a popular approach to stimulate students and create a more positive classroom environment. It can serve as an alternative or supplement to traditional teaching and can better engage students while developing a positive attitude toward learning. The use of games in a classroom can create a more exciting and engaging environment, while still reinforcing learning concepts. This paper explores the development of three math games to supplement classroom learning and analyses the game mechanics and designs implemented in each game.

Keywords: Game-based learning, Mathematics, Teaching Supplement

Education & Training

Promoting Skills in Children and Teens with Autism Spectrum Disorder Through Play and STEAM

Meaghan McLeod Mazingo, Krzysztof Rechowicz

Individuals with autism spectrum disorder (ASD) have a low employment rate. This is caused by a lack of support from employment resources and the negative stigma associated with common characteristics associated with ASD. With limited career-building activities and events, it is difficult for individuals with ASD to identify their skills, strengths, and career opportunities. Parental support is crucial when seeking employment opportunities for their children. Through the use of play, children with ASD engaged and explored their skills with science, technology, engineering, arts, and math-centered activities. This paper highlights the events and shows the proposed redesign for an additional workshop.

Keywords: Autism spectrum disorder, STEAM, Workforce development, Inclusion, Accessibility
