Machine Learning for Behind Armor Blunt Trauma Risk Assessment: Improving Usability in I-PREDICT (Incapacitation Prediction for Readiness in Expeditionary Domains: An Integrated Computational Tool)

Vivek B. Kote, Lance L. Frazer, Koen Flores, Brian Connolly, Tim Bentley, Daniel P. Nicolella

I. INTRODUCTION

Finite element (FE) human body models (HBMs) have become a powerful tool for simulating the body's response to dynamic impacts, offering a cost-effective and adaptable alternative to traditional experimental methods. However, conventional injury risk assessment approaches with FE models rely on average 50th percentile male models, overlooking variability in tissue properties and anatomical morphology [1-3]. To address this, the DoD-funded I-PREDICT Future Naval Capability (FNC) developed a probabilistic HBM framework that incorporates these variations, enabling population-specific injury risk assessments [4]. Recently, the I-PREDICT HBM was used to assess injury risk for behind armor blunt trauma (BABT). Despite their accuracy, such high-fidelity models are computationally expensive and require specialised expertise, limiting their usability. To address this, machine learning surrogate models were trained on strain output from each simulation, enabling rapid strain predictions and injury risk estimation without the computational burden of full FE simulations [5-6]. Furthermore, we explored the integration of Large Language Models (LLMs) to create an intuitive interface for I-PREDICT. The LLM-powered interface streamlines data processing, organises model inputs, and samples surrogate models, enabling users to assess injury risks for critical regions without requiring in-depth knowledge of FE modeling. This study focuses on integrating machine learning methods, with a particular emphasis on the use of LLMs, to enhance the usability and practicality of I-PREDICT, thereby making advanced injury prediction more accessible for operational and research applications.

II. METHODS

Computational Modeling

FE simulations in LS-DYNA were conducted using impactor specifications and velocities designed to mimic the typical backface armour deformation in accordance with the NIJ 0101.06 standard [7]. The I-PREDICT HBM incorporated variations in impact location (±20 mm) and 27 tissue properties for organs of interest from literature. A total of 450 simulations were performed for each impact site – heart, liver, and lower abdomen (Fig. 1) – using uniform distributions for variables in the model describing material properties, body size (discrete uniform distribution - 1.70m, 63.5kg to 1.87m, 79.3kg), impact speed (20m/s to 53m/s) and location [5].

The maximum principal strain across all time points for each element in organs of interest (liver, kidney, ribs, heart, lung, spleen) was used to train Gaussian Process surrogate models for injury prediction. To improve efficiency, Principal Component Analysis (PCA) was applied to the strain data, reducing the number of variables by predicting only the principal component (PC) weights instead of strain values for each element. The full strain field was reconstructed using inverse PCA, with the number of PCs chosen to retain at least 95% of strain variability. Model validation was performed using leave-one-out analysis [6]. To estimate population-level injury risk, the surrogate models were sampled 1,000 times using Latin Hypercube Sampling (LHS) using appropriate distributions from literature. Injury probabilities were reported using the Military Combat Injury Scale (MCIS), and an overall incapacitation risk was determined via the New Injury Severity Score (NISS) [5].

To enhance usability, an 'Injury Assistant' (IA) tool was integrated with I-PREDICT, leveraging four task-specific LLMs: Sensical, Input Parser, Converter, and MCIS (Fig. 2). The task-specific LLMs were implemented as instances of GPT-4o hosted in a FedRAMP-compliant environment. The Sensical LLM first identifies whether a user's query matches a relevant scenario. If so, the Input Parser LLM gathers necessary information interactively while filtering out irrelevant details. The Converter LLM processes this input to execute pre-trained surrogate models, generating injury predictions. Finally, the MCIS LLM translates the injury risk results into conversational terms.

The IA tool was tested by 21 novice users, with logs tracking successful queries and corresponding injury predictions. To verify its accuracy, expert users independently assessed the same prompts and compared their manually derived predictions with the IA's outputs to evaluate its success rate.

V. B. Kote (e-mail: vivek.kote@swri.org; tel: +12105222528), L. L. Frazer, K. Flores, B. Connolly, and D. P. Nicolella work at Southwest Research Institute in San Antonio, Texas, USA. T. Bentley works at the Office of Naval Research, Arlington, Virginia, USA.

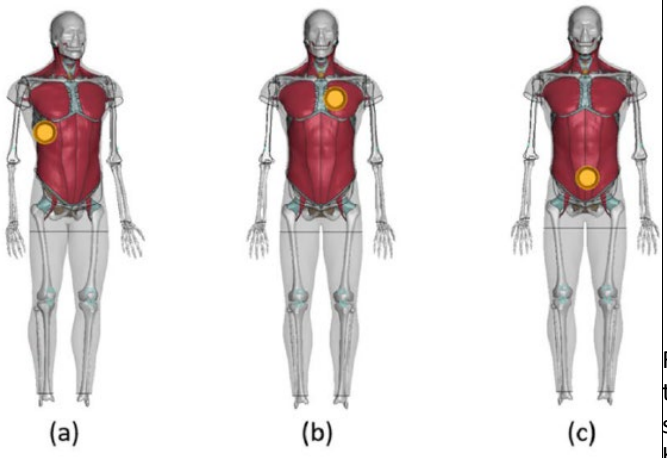

Fig. 1. Three impact sites tested: (a) liver, (b) heart, and (c) lower abdomen.

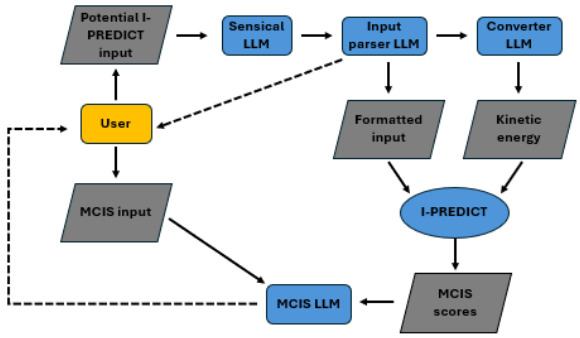

Fig. 2. Flowchart illustrates interactions between the LLM and the user. Dotted lines denote the conversation loop, while solid lines indicate data flow. Blue boxes represent processes handled by the LLM, and the yellow box highlights areas requiring user input. The blue circle signifies the call to the I-PREDICT response surfaces, with grey elements representing data components.

III. INITIAL FINDINGS

FE simulations were conducted using eight LS-DYNA cores, with each run, including post-processing, taking approximately 18 hours. The number of PCA modes required to train the response surface varied by impact location, with organs closer to the impact site exhibiting greater variability and therefore requiring more modes to accurately capture strain distribution. For example, the spleen required six modes to capture 95% variability for liver impacts, while the 6th rib required 24 modes. Scree plots indicated that the first PCA mode explained ≥85% of the strain data variability. Leave-one-out validation of the surrogate models produced R^2 values between 0.90 and 0.99 for the first mode shape, meeting the predictive accuracy threshold (cumulative $R^2 \geq 0.85$). Predictive accuracy decreased with subsequent modes based on variance explained. Once verified, the surrogate model enabled whole-body injury risk predictions using 1000 Latin Hypercube samples in approximately 15 minutes [5]. The LLM-powered IA tool was also evaluated for accuracy. It successfully extracted all relevant user information in conversational terms and was full consistent with injury risk predictions (Fig. 3) from expert I-PREDICT users, demonstrating its effectiveness in streamlining complex computational processes.

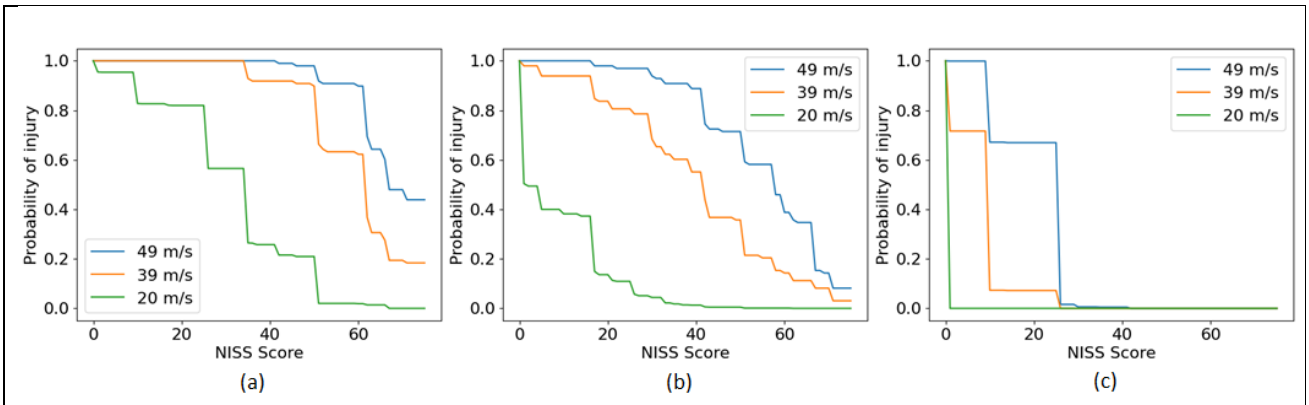

Fig. 3. NISS-based injury risk curves showing the probability of sustaining an injury with a NISS score equal to or greater than a given value, for three different impact velocities, at (a) the heart, (b) the liver, and (c) the lower.

IV. DISCUSSION

Results from this study indicate that current body armour protection standards might result in severe injuries in certain locations, whereas no injuries in others. Additionally, this study found that machine learning techniques, such as surrogate models and LLMs, greatly reduce computational time and significantly shorten the training period for new users. The adaptable framework effectively bridges knowledge gaps, enabling novice users to interact with complex models like I-PREDICT without extensive prior expertise. Future efforts will focus on expanding the surrogate model dataset to cover a broader range of BABT scenarios and other relevant applications, further improving the IA's accuracy and versatility.

V. REFERENCES

[1] Iraeus, J., *et al*., *J Mech Behav Biomed Mater*, 2020. [2] Nicolella, D. P., *et al*., *J Biomech*, 2006. [3] Nicolella, D. P., *et al*., ASME BED, 2001. [4] DiSerafino, *et al.*, ABME, 2024. [5] Kote, V. B., *et al.*, ABME, 2024. [6] Frazer, L., *et al.*, CMBBE, 2023. [7] Op't Eynde, *et al*., PASS, 2020.