medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

A Real-Time Deep Learning Approach for Inferring Intracranial Pressure from Routinely Measured Extracranial Waveforms in the Intensive Care Unit

Shiker S. Nair,[1] Alina Guo,[1] Joseph Boen,[1] Ataes Aggarwal,[1] Ojas Chahal,[1] Arushi Tandon,[1] Meer Patel,[1] Sreenidhi Sankararaman,[1] Tej Azad,[2] Romain Pirracchio,[3] and Robert D. Stevens[1,2,4]

[1] *Department of Biomedical Engineering, Johns Hopkins University Whiting School of Engineering, Baltimore, MD, USA*
[2] *Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
[3] *Department of Anaesthesia and Perioperative Care, UCSF, San Francisco, CA, USA*
[4] *Departments of Anesthesiology and Critical Care Medicine, Neurology, and Radiology, Johns Hopkins School of Medicine, Baltimore, MD, USA*

Abstract - Objective. Intracranial pressure (ICP) is a physiological variable used to assess the neurological state of patients with life-threatening intracranial pathology, such as traumatic brain injury or stroke. The current standard of care for measuring ICP requires a catheter to be inserted into the brain, which is associated with an appreciable risk of hemorrhage and infection. We hypothesize that ICP can be computed from extracranial waveforms routinely measured in the Intensive Care Unit (ICU), such as invasive arterial blood pressure (ABP), photoplethysmography (PPG), and electrocardiography (ECG). **Methods.** We extracted 600 hours of simultaneous ABP, ECG, PPG, and ICP data (sampled at 125 Hz) across 10 different patients from the MIMIC III Waveform Database. These recordings were segmented into 10 second windows and used to train six different deep learning models with ABP, ECG, and PPG waveforms as input features. Models were evaluated in both a single-patient analysis and multi-patient analysis. **Results.** The performances of the six deep learning models were compared, revealing two tiers of performance. Among the top-tier models, the mean average error (MAE) for inferring ICP was approximately 1.50 mmHg for single-patient analysis and 5 mmHg for multi-patient analysis. **Conclusions.** These preliminary and novel results indicate the feasibility and accuracy of noninvasive ICP estimation by training deep learning models with extracranial physiological data. With further validation, this approach could be implemented in a continuous real-time fashion, thereby reducing risks associated with invasive monitoring and allowing more timely treatment of patients with critical brain injuries.

Index Terms— Arterial blood pressure, deep learning, electrocardiogram, intracranial pressure, noninvasive diagnosis, photoplethysmography, precision medicine

I. Introduction

Traumatic brain injury (TBI), a type of brain damage caused by physical trauma to the head, is one of the leading causes of death and disability worldwide [1]. In the US alone, recent data indicate 223,135 TBI-related hospitalizations and close to 70,000 TBI-related deaths [2]. The leading cause of TBI-related death is increased intracranial pressure (ICP) [3]. ICP is defined as the pressure exerted within the cranial compartment of a patient [4]. At rest, a normal supine adult's ICP is 0-20 mmHg [5,6]. Intracranial hypertension is defined as a sustained increase of ICP above 20 mmHg [4]. This is typically caused by an expanding intracranial mass, cerebral edema, or a process impeding cerebrospinal fluid flow and/or reabsorption through the ventricular and subarachnoid spaces [7].

ICP monitoring is highly recommended in patients with severe TBI [3]. In cases of elevated ICP, immediate treatment is required for the patient (i.e., administration of hypersaline agents) [4]. Currently, the standard for monitoring ICP is the placement of an external ventricular drain (EVD) or intraparenchymal brain monitor (IPM) [5,8]. Insertion of such devices requires a hole to be drilled in the skull and the probe to be inserted through the tissues of the brain. The invasive nature of this procedure poses risks including infection and hemorrhage. The risk of infection for the EVD is 5-14% and the risk of hemorrhage is 5-7% [3,6]. Additionally, insertion of an EVD requires a neurosurgery team and takes approximately 30-60 minutes, resulting in delayed ICP measurements potentially delaying ICP detection and intervention [9]. This could result in rapid deterioration of a TBI patient and, eventually, lead to death [4]. Thus, there remains a need for a less invasive, continuous way of monitoring ICP that is reliable and accurate.

Previous research has indicated coupling between ICP and the cardiovascular system, suggesting that information present in extracranial physiological waveforms is correlated with intracranial events [10]-[12]. Cerebral autoregulation may be impaired as a consequence of elevated ICP, leading to pathological changes in cerebrovascular vasoreactivity, cerebral blood flow, and oxygenation. Here we explore the relationship between ICP and three physiological indices which are routinely measured/monitored in the ICU: invasive arterial blood pressure monitoring (ABP), photoplethysmography (PPG), and electrocardiography (ECG). We hypothesize that an estimated ICP measure can be computed by decoding extracranial physiological waveform data.

Estimating ICP using extracranial waveforms would allow clinicians to assess patients with intracranial abnormalities in a noninvasive, cost-effective manner. Furthermore, such a solution could be unobtrusively integrated into the current clinical workflow in the ICU.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

II. Related Works

Our research presents a novel approach to predicting ICP in patients with TBI that significantly differs from existing studies in the literature. We leverage traditional machine learning and deep learning architectures to predict ICP using high-frequency (125 Hz) ABP, PPG, and ECG waveform data in real-time. Although previous groups have attempted to develop both physics-based simulations and statistical models to infer ICP, there have been limited efforts to use data-driven models to accomplish this task [13, 14].

The primary novelty of our approach lies in the use of solely extracranial signals to predict the ICP waveform in quasi-real-time, using deep learning techniques. Similar studies in this field have successfully developed methods to predict the occurrence of hypertensive events or the progression of intracranial hypertension (IH) into a potentially life-threatening condition [15]-[18]. Other works have developed deep-learning algorithms to forecast the ICP waveform [19,20]. However, all of these studies rely on using ICP as an input and therefore do not offer the potential of an extracranial, non-invasive methodology for measuring ICP. There is indeed literature for the real-time prediction of ICP using extracranial signals [13,14]; however, these studies scarcely utilize deep learning in their formula-based methods.

By focusing on non-invasive methods for measuring ICP, our research has the potential for transforming the monitoring and treatment of TBI patients, particularly in resource-limited settings. ICP could become a routinely estimated vital sign beyond the intensive care unit (ICU). In addition, real-time waveform re-creation using deep learning can be seamlessly integrated into the current monitoring workflow to replace invasive catheters while still giving clinicians the autonomy to interpret ICP waveforms.

III. Data

MIMIC-III (v1.4, 2016) is an intensive care database that comprises de-identified data from 46,520 patients and 58,976 admissions, recorded between June 1, 2001, and October 31, 2012, at the Beth Israel Deaconess Medical Center in Boston, USA [21]. The data includes demographics, admission, and discharge notes, International Classification of Diseases - 9th revision (ICD-9) codes for diagnoses, radiological findings, survival data, vital sign recordings, medications, and procedures for each patient [21]. MIMIC-III also contains high-frequency physiological waveforms data from a subset of patients. We decided to model ICP from ABP, ECG, and PPG. By choosing these waveforms, we had three modalities of predicting ICP and assessing the contributions of each waveform in the modeling tasks.

IV. Methodology

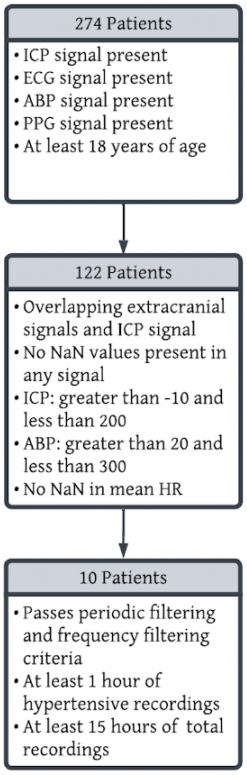
A. Preprocessing, Cohort Selection, and Patient Demographics

Patients who were at least 18 years old with ECG, PPG, ABP, and ICP recordings for at least five minutes were included in this study. To accomplish this, we first assessed whether a patient had recordings of all four waveforms. Because ECG is measured at multiple different leads, we used the most prevalent ECG lead (II) in our patient population of interest. The rationale behind the common usage of ECG lead II for ICP prediction is rooted in its ability to provide a reliable depiction of the mechanical forces engendered by the cardiac system and their subsequent impact on ICP [22]. After identifying patients who have ICP, ABP, ECG Lead II, and PPG data available, we removed patients who had less than five minutes of overlapping data between these four waveforms. This was done based on the knowledge that elevated ICP lasting for five minutes or more may indicate a hypertensive event, according to clinical experts [23]. Therefore, it was important to ensure that we had sufficient data to accurately detect such an event if it occurred.

From the patients that remained, the waveform data (synchronized in time) were sliced into five-minute segments. To ensure data quality, five-minute segments that contained any not a number (NaN) values or physiologically unrealistic values (ICP $<$ -10 mmHg or $>$ 200 mmHg, ABP $<$ 20 mmHg or $>$ 300 mmHg, not NaN mean heart rate) unique to each waveform were filtered out. To address potential issues with EVD calibration, an additional periodicity filtering step was applied to the ICP signals. This filtering step removed extremely high-frequency and low-frequency waveforms that could lead to inaccuracies in the data. To ensure that intracranial hypertension events were adequately represented in the data, patients who had less than one hour of intracranial hypertension recordings were excluded. For this study, an ICP hypertensive event was defined as an ICP greater than 20 mmHg sustained for at least five minutes [23]. Although recent traumatic brain injury guidelines propose a threshold of 22 mmHg, our study includes a heterogeneous population of patients who underwent ICP monitoring for both traumatic and nontraumatic brain disorders [4]. Therefore, we opted to use a threshold of 20 mmHg. Figure 1 summarizes these pre-processing workflow for cohort selection.

After applying this pre-processing scheme, 10 patients remained for downstream analysis containing more than 600 hours of recorded ICP, ABP, EKG, PPG data (more than 270,000,000 data points). Table I highlights key demographics features of these patients.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .




```
Result: Final list of patients that meet all inclusion/exclusion criteria
patients ← dataset of 274 patients with ICP, ECG, PPG, and ABP
  present and at least 18 years of age;
filtered_patients ← empty list;
for patient in patients do
    icp_signal ← get_icp_signal(patient);
    extracranial_signals ← get_extracranial_signals(patient);
    overlapping_signals ← get_overlapping_signals(icp_signal,
     extracranial_signals);
    if len(overlapping_signals) < 5 minutes then
        continue;
    end
    if any(signal_value is NaN for signal_value in overlapping_signals)
     then
        continue;
    end
    if any(signal_value < -10 or signal_value > 200 for signal_value in
     icp_signal) then
        continue;
    end
    abp_signal ← get_abp_signal(patient);
    if any(signal_value < 20 or signal_value > 300 for signal_value in
     abp_signal) then
        continue;
    end
    hr_signal ← get_hr_signal(patient);
    if any(signal_value is NaN for signal_value in hr_signal) then
        continue;
    end
    filtered_patients.append(patient);
end
hypertensive_patients ← empty list;
for patient in filtered_patients do
    icp_signal ← get_icp_signal(patient);
    if mean(icp_signal) ≤ 20 then
        continue;
    end
    if total_recording_hours(patient) < 15 then
        continue;
    end
    if hypertensive_recording_hours(patient) < 1 then
        continue;
    end
    hypertensive_patients.append(patient);
end
periodic_and_frequency_filtered_patients ← empty list;
for patient in hypertensive_patients do
    if at_least_one_hour_hypertensive_recording(patient) and
     at_least_15_hours_total_recording(patient) and
     frequency_filtering(patient) then
        periodic_and_frequency_filtered_patients.append(patient);
    end
end
final_patients ← empty list;
for patient in periodic_and_frequency_filtered_patients do
    final_patients.append(patient);
end
return final_patients;
```

Algorithm 1: Inclusion/exclusion criteria for patient selection

FIG. 1: Cohort Selection Workflow (Left) and Pre-Processing Methodology (Right)

Table I. Patient Demographic Data in Study

Total Number of Subjects	n = 10
Age Median [IQR]	56 [33-64]
Sex(% M, % F)	30, 70
Admission Type (Number of Subjects)	Emergency (8) Urgent (2)
Discharge Location (Number of Subjects)	Rehab/Distinct Part Hosp (5) Dead/Expired (3) Home Health Care (1) Home (1)
Diagnosis (Number of Subjects)	Subarachnoid hemorrhage (5) Acute liver failure (2) Traumatic brain injury (1) Intracerebral hemorrhage (1) Closed fracture of base of skull (1)

B. Determining Window Length for ICP Prediction Task

To ensure that the ICP prediction task was as close to real-time as possible, a window size of 10 seconds was chosen. To assess the amount of information lost as the window size decreased, a dynamic time warping analysis (DTW) was conducted [24]. The patient data was divided into windows of 10 seconds, 20 seconds, 30 seconds, and 60 seconds. The DTW score for each of the extracranial waveforms (ABP, ECG, and PPG) and ICP was computed between adjacent windows. This resulted in four scores (one for each waveform) for each pair of adjacent windows. The Spearman correlation between the DTW scores for each extracranial waveformand ICP was calculated (ABP/ICP, ECG/ICP, and PPG/ICP) to assess if changes in the extracranial waveforms would be reflected in ICP.

The results showed that all window sizes had a weak positive correlation between the extracranial waveforms and ICP (Table II). This justified the choice of a 10-second window size to best replicate real-time analysis.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Table II. Dynamic Time Warping Analysis for Assessing Information Loss Across Decreasing Window Sizes

Window Size (s)	ICP/ABP	ICP/PPG	ICP/ECG
10	0.39	0.29	0.25
20	0.37	0.30	0.29
30	0.40	0.33	0.31
60	0.42	0.37	0.35

C. Model Selection and Architecture

The goal of this study is to predict the entire ICP waveform (125 Hz) from ABP, ECG, and PPG waveforms (125 Hz) in 10-second windows. Given this is a multivariate time series prediction problem, we implemented six different deep learning models which are fit to accomplish this task and have been used in previous studies in similar problem spaces for time-series based predictions (Figure 2):

- Recurrent Neural Network (RNN) [25,26]
- Gated Recurrent Unit (GRU) [27]-[30]
- Long-Short Term Memory (LSTM) network [19], [30- 32]
- Temporal Convolutional Networks (TCN) [33]-[36]
- Volume-Net (VNET) [37]
- Transformer [38,39]

These six models vary in sophistication and complexity. In general, adding more layers improved the predictions for each model, however, this increased the risk of overfitting. Rigorous hyperparameter tuning was applied for each model, specifically finding an optimal learning rate, batch size, and number of nodes for each layer.

D. Normalization and Evaluation Criteria

Both features (ABP, ECG, and PPG) and labels (ICP) waveforms were independently normalized for the prediction tasks. Normalization was completed using a standard scaling function (subtracting by mean and dividing by the standard deviation of the waveform) which was applied to all the data before splitting it into a training and testing set. While it is common practice to normalize the training set and apply this normalization scheme to the test set, the small number of patients used in this analysis (n=10) did not allow for a generalizable normalization function in the training set that would properly transfer over to the testing set. As a result, all the data was normalized before splitting into a training and testing set. Predictions were then unscaled back into a clinically interpretable ICP using the inverse scaling function.

Different loss functions were assessed when optimizing each model and it was determined that the log hyperbolic cosine (log cosh) loss was the best optimization function based on convergence. It is defined as the logarithm of the hyperbolic cosine of the difference between the true (t) and predicted values (p) (Equation 1). The logcosh loss function is a smooth approximation of the MAE loss function, allowing it to be easily optimized through gradient descent methods. Furthermore, logcosh is known to converge faster than other loss functions such as MAE, overall making it an ideal function for deep learning tasks [40]-[42]. MAE and root mean squared error (RMSE) were also metrics used to report model performance (Equation 2 and Equation 3).

Equation 1. Logcosh Loss Function Formula

$$Logcosh(t) = \sum_{p \in P} log(cosh(p - t))$$

Equation 2. Mean Average Error Function Formula

$$MAE = \frac{1}{n} \sum_{p \in P} |p - t|$$

Equation 3. Root Mean Squared Error Formula

$$RMSE = \sqrt{\sum_{p \in P} \frac{(p - t)^2}{n}}$$

E. Single-Patient Analysis

The first experiment conducted evaluated the performance of each of the six models within a single-patient for each of the 10 patients in the study. The goal of this experiment was to ascertain that each model was capable of predicting ICP waveforms from the extracranial waveforms in the simplest context: within a single-patient. This removes any notion of generalizability and enables the comparison of each model in a simpler context before considering inter-patient generalization. Data were split into 70% train/validation set and 30% test set. The mean and median logcosh for training and testing datasets were reported for each model along with the standard deviation. Using this model information, we selected the top model(s) to run a multi-patient analysis. Additional follow-up analysis was performed to understand the strengths and weaknesses of each model and trends in error and prediction.

F. Multi-Patient Analysis on Top Models

To ensure that the top models are generalizable to new patients, a multi-patient analysis was conducted using 6

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

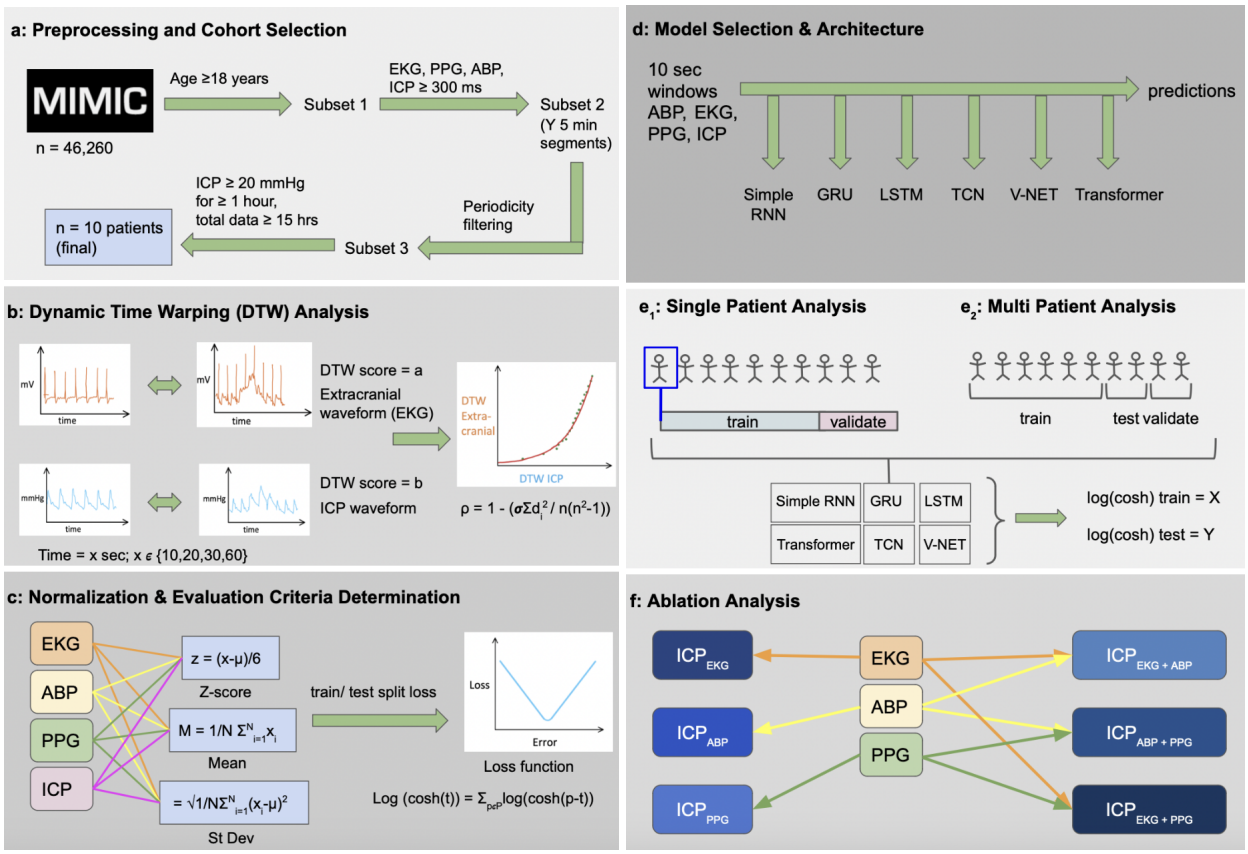


FIG. 2: Overview of Methodology and Analysis Performed in Study

patients for training, 2 for validation, and 2 for the test set. This approach ascertains that the models are evaluated on data from patients that are distinct from those used in the training set, thus providing a robust assessment of model performance. The mean and median log-cosh values for the training and testing datasets were reported for each model, to provide a measure of model performance accuracy and consistency. Furthermore, additional follow-up analyses were performed to understand trends across the top models and across different patients in the validation and testing sets. This analysis is crucial for ensuring that the models are generalizable and can be applied to new patients in clinical practice, where the algorithm will be trained on different patients than the patients it will be used on in the ICU.

G. Ablation Analysis on Multi-Patient Top Model

To assess extracranial waveform importance, an ablation study was conducted in the top model for multi-patient analysis. All subsets of input extracranial waveforms were tested: ABP only, PPG only, ECG only, ABP + PPG, PPG + ECG, and ABP + ECG. The goal of this study was to understand the effect that dropping extracranial waveforms had on model performance.

H. Clinical Error Analysis Procedure

To improve the accuracy of predictions for individual patients, an error analysis was conducted to identify the factors contributing to the varying degrees of loss across the testing dataset. The analysis investigated potential correlations between loss and various factors, including autoregulation index, mean ICP values, ICP signal variance, and clustering of features in a reduced dimensionality space. This investigation aimed to identify any underlying reasons for the variation in loss across the testing dataset and to inform the development of improved models for predicting ICP in individual patients and across patients.

To investigate the relationship between the autoregulation index and loss, we computed the Pearson correlation coefficient between the ABP and ICP to obtain an autoregulation index for each prediction in the testing set for every 10 second window instance. The autoregulation index represents the degree of correspondence between ABP and ICP waveforms. We hypothesized that a higher autoregulation index would be associated with lower loss.

Next, data imbalance was assessed to examine whether poorly represented data resulted in worse predictions. This was particularly important in cases of high mean

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

ICP, which was a small fraction of all data yet the most clinically relevant. To assess whether a high ICP was associated with poor prediction, the mean ICP value was calculated for each window of the testing dataset, and the Pearson correlation coefficient was computed between the ICP value and loss.

Additionally, to investigate the correlation between ICP signal variance and loss, the variance of the ICP signal was calculated for each instance, and the Pearson correlation coefficient was computed between the loss and the ICP signal variance.

Lastly, to examine if the quality or similarity of input features influenced loss, all three input waveforms were reduced into a two-dimensional space using the Uniform Manifold Approximation and Projection (UMAP) Python library. loss was overlaid onto these points to assess cluster-specific trends.

V. Results

In the single-patient analysis for predicting ICP using ABP, PPG, and ECG in 10-second windows across 10 patients, the average MAE testing loss for Simple RNN was 3.17±0.68 mmHg, GRU was 1.31±0.59 mmHg, LSTM was 1.63±0.60 mmHg, TCN was 1.18±0.57 mmHg, V-NET 1.22±0.57 mmHg, and Transformer was 2.65±0.82 mmHg (Table III). Model performance on for each patient can be found in Supplementary Tables 1A-1F. Based on these results, there were two tiers of model performance. GRU, LSTM, TCN, and V-NET all performed significantly better than the Simple RNN and Transformer models (refer to Supplemental Table 2A and 2B for significance testing). While there is some variation in results for GRU, LSTM, TCN, and V-NET, there is no conclusive 'top' model.

The multi-patient analysis results show that all models perform to a similar degree with an MAE of approximately 5 mmHg (Table IV). Figure 3 shows a sample ICP waveform prediction by the TCN model. Figure 4 shows the Bland-Altman graphs for each model in the multi-patient analysis with a testing set mean difference of 0.56 mmHg for Simple RNN, -0.26 mmHg for GRU, 0.09 mmHg for LSTM, 0.84 mmHg for TCN, 0.81 mmHg for V-NET, and 1.53 mmHg for Transformer. Upon further inspection of the Bland-Altman graphs for each model, it becomes clear that there is variation in model predictions. Both the Simple RNN and Transformer models have a semi-linear trend in the Bland-Altman graphs, indicating the presence of a systematic bias. In contrast, GRU, LSTM, TCN, and V-NET have no decisive correlation across the mean difference line. This trend is also observed in testing set predictions versus ground truth labels for each model (Supplementary Figure 1). There are a few lines of reasoning that may help to explain the discrepancies in model performance for the superiority of LSTM, GRU, TCN, and V-NET over the Simple RNN and Transformer. The first is the data representation itself. Transformers are particularly effective when there is a clear spatial structure in the data, such as in image or language processing tasks. They excel at capturing global dependencies but may struggle with capturing local and temporal dependencies, which are prevalent in waveform data. Similarly, Simple RNNs might have difficulty capturing long-term dependencies due to the vanishing gradient problem. On the other hand, TCN, GRU, LSTM, and V-NET architectures are designed to address these issues, making them more suitable for the task at hand with the given structure of the data. In terms of model capacity, TCN, GRU, LSTM, and V-NET architectures generally have larger model capacities compared to Simple RNN and Transformer architectures in this case. The larger capacity allows them to learn more intricate patterns and relationships in the data, which may be necessary for accurately inferring the ICP waveform. Therefore, having such discrepancies across models may elucidate insights into the nature of the computations necessary for accurate inferences.

The ablation study for multi-patient analysis demonstrated that all three waveforms are necessary for producing the most optimal ICP prediction (Table V). When only inputting one waveform, using ECG yielded the best results across almost all models followed by PPG and ABP. When using a pair of extracranial waveforms, there was no clear trend in an optimal combination. The outlier model in the ablation study was V-NET, which showed only a small improvement in predictions with the addition of multiple extracranial waveforms. This exception to our expectations may again offer valuable insight into model architectures that are most conducive to performing accurate inferences. The V-NET architecture incorporates mechanisms and layers that enable dimensionality reduction, allowing it to capture essential information effectively. Utilizing pooling layers, the V-NET reduces spatial dimensions by downsampling feature maps. Additionally, the Conv1D layers perform 1D convolutions, capturing local patterns and dependencies, effectively reducing dimensionality and extracting relevant features, including important temporal dependencies.

Finally, in conducting our clinical error analysis, we found that model performance was not significantly correlated with autoregulation index, mean ICP values, ICP signal variance, or clustering of features in a reduced dimensionality space. Ultimately, further investigation is required to determine the causal factor for poor loss in our algorithms.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Table III. Summary of Single Patient Analysis Results

	Simple RNN	**GRU**	**LSTM**	**TCN**	**V-NET**	**Transformer**
Logcosh Test Loss (Train Loss) in mmHg	2.57 ±0.66 (2.69 ± 0.73)	0.86 ±0.52 (0.73 ± 0.45)	1.08 ±0.54 (0.91 ± 0.51)	0.75 ±0.50 (0.41 ± 0.26)	0.80 ±0.50 (0.39 ± 0.32)	2.08 ±0.79 (2.15 ± 0.79)
MAE Test Loss (Train Loss) in mmHg	3.17 ±0.68 (3.36 ± 0.78)	1.31 ±0.59 (1.13 ± 0.52)	1.63 ±0.60 (1.38 ± 0.62)	1.18 ±0.57 (0.64 ± 0.30)	1.22 ±0.57 (0.58 ± 0.39)	2.65 ±0.82 (2.74 ± 0.84)
RMSE Test Loss (Train Loss) in mmHg	4.26 ±0.81 (4.52 ± 0.94)	1.95 ±0.77 (1.69 ± 0.70)	2.26 ±0.89 (1.87 ± 0.84)	1.83 ±0.78 (1.02 ± 0.41)	1.86 ±0.77 (0.89 ± 0.52)	3.61 ±1.07 (3.73 ± 1.12)

Table IV. Multi-Patient Analysis Results

	Simple RNN	**GRU**	**LSTM**	**TCN**	**V-NET**	**Transformer**
Logcosh Test Loss (Train Loss) in mmHg	4.38 (1.94)	4.55 (2.03)	4.56 (2.01)	4.33 (1.89)	4.41 (1.95)	5.10 (2.43)
MAE Test Loss (Train Loss) in mmHg	5.02 (2.41)	5.19 (2.70)	5.20 (2.47)	4.98 (2.45)	5.05 (2.50)	5.75 (2.94)
RMSE Test Loss (Train Loss) in mmHg	6.25 (3.65)	6.47 (3.89)	6.48 (3.74)	6.11 (3.63)	6.23 (3.68)	7.00 (4.21)

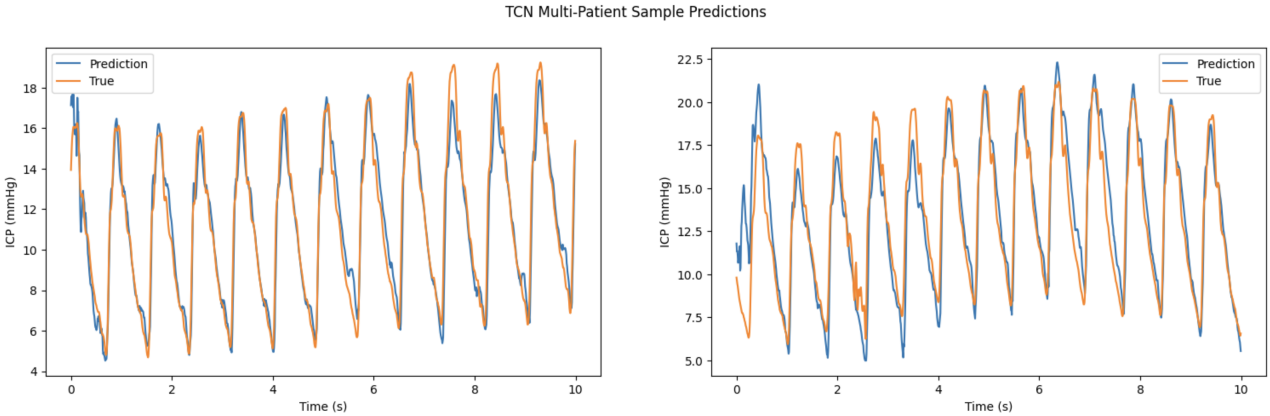


FIG. 3: TCN Multi-Patient Waveform Prediction Visualizations

VI. Discussion

A. Key Results

Using extracranial continuous physiological waveform data, we present a non-invasive modeling approach for computing a surrogate ICP from extracranial waveforms across six different deep learning models. Our results suggest that deep learning algorithms can model ICP from extracranial waveforms as demonstrated by our single-patient analysis where top models (GRU, LSTM, TCN, V-NET) predicted ICP with a MAE of approximately 1.50 mmHg. Furthermore our results show our algorithms are proficient in generalizing across patients in a small training set with our top models predicting ICP with a MAE of approximately 5 mmHg with limited systematic bias. Beyond this, we demonstrate that PPG is an important waveform in predicting ICP, a novel insight. The strength of these results can be partly explained by the relationship between these extracranial waveforms and intracranial hemodynamics. In addition, these results show there exists valuable information encoded in the waveform morphologies of ECG, PPG, and ABP that relate to ICP.

There is only one study to our knowledge that aims at deriving ICP non-invasively from waveform data using transcranial doppler (TCD) [45]. This study used ABP, ECG, and cerebral blood flow velocity (derived

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

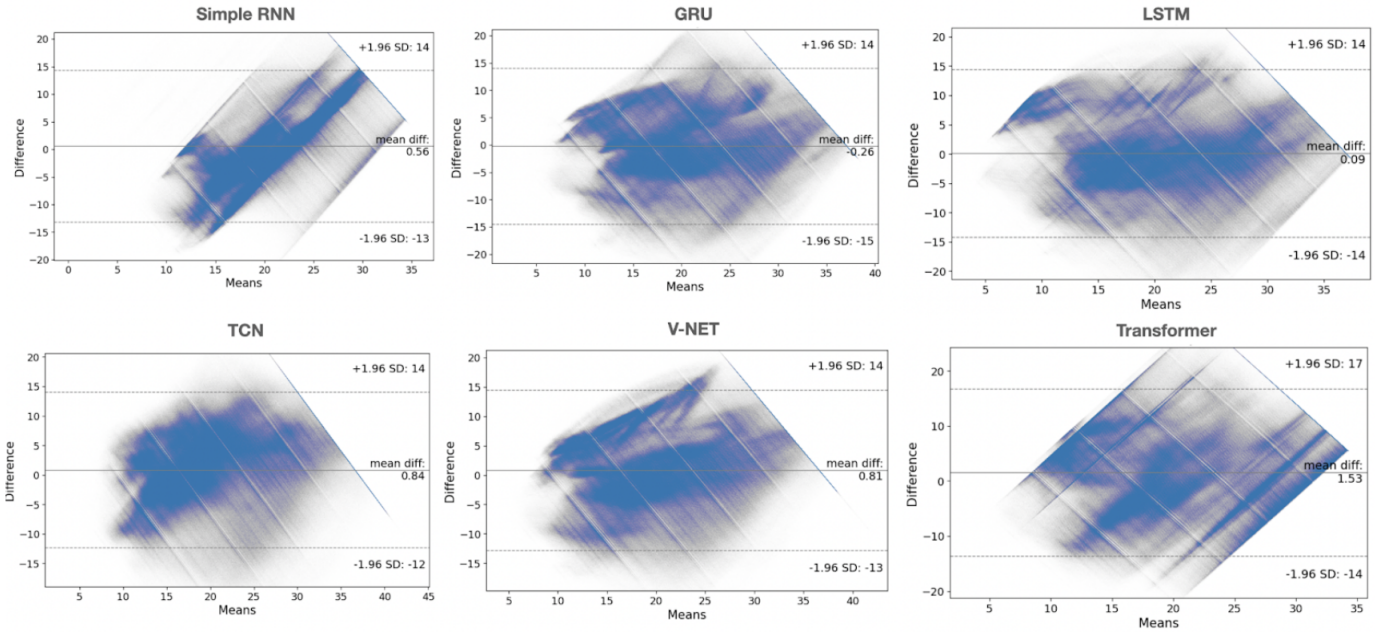


FIG. 4: Bland-Altman Graphs for Multi-Patient Analysis for All Six Models

Table V. Ablation Results for TCN Multi-Patient Analysis

Ablation Test	RNN Test Set MAE Loss (mmHg)	GRU Test Set MAE Loss (mmHg)	LSTM Test Set MAE Loss (mmHg)	TCN Test Set MAE Loss (mmHg)	VNET Test Set MAE Loss (mmHg)	Transformer Test Set MAE Loss (mmHg)
ABP	8.23	7.51	8.14	8.11	5.42	8.46
PPG	6.83	7.04	7.20	7.13	5.21	8.34
ECG	6.57	6.79	6.72	6.54	5.24	8.31
ABP + PPG	5.23	6.32	5.78	5.45	5.17	7.44
ABP + ECG	5.91	5.78	5.66	5.36	5.32	6.33
PPG + ECG	5.12	5.24	6.18	6.03	5.06	6.95
ABP + PPG + ECG	5.02	5.19	5.20	4.98	5.05	5.75

from TCD) as input waveforms for domain adversarial neural networks and transformers to infer ICP across 11 patients with a median MAE of 3.88+/-3.26 mmHg and 3.94+/-1.71 mmHg, respectively. A limitation with this study is that it uses cerebral blood flow velocity, which is not a routinely measured waveform in the ICU given that its computation requires a TCD ultrasonography scan. As such, this data study only uses around 500,000 data points in comparison to our study which uses more than 270,000,000 data points.

We can also compare our algorithms' performances to other on-the-market technologies that aim to measure ICP both less-invasively and invasively. TCD ultrasonography has an error range of ± 12 mmHg [46] and cannot monitor ICP continuously. Tympanic membrane displacement has an error range of ± 15 mmHg and requires a probe to be inserted into the canal of the ear [8]. Optic sheath diameter, a retinal imaging-based technique, reports a margin of error from 5-10 mmHg and also cannot monitor ICP continuously [8]. Furthermore, our top algorithms performed similarly to other state-of-the-art alternatives to the EVD such as intracranial transducers and intraparenchymal monitoring which both report a mean error of ± 7 mmHg [8,46]. While both techniques offer real-time and continuous monitoring, they are on the same level of invasiveness as an EVD, requiring surgical implantation.

B. Novelty

This research provides compelling evidence that it is possible to accurately and in real-time compute a surrogate ICP from extracranial waveforms. With further validation, the real-time and non-invasive nature of our algorithms suggest they could be integrated into the current clinical workflow of ICP monitoring, a profound advancement towards eliminating the need for an EVD. In

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

addition, by assessing multiple deep learning algorithms for accomplishing this task, we have also conducted a robust methodological study on the performance of deep learning algorithms on time-series data. Understanding the differences in predictions from these six deep learning algorithms could provide insights into understanding both the clinical and technical components of this multifaceted problem. Ultimately, the results of this study demonstrate that algorithmic prediction of ICP performs better than current, less invasive alternatives and on par with alternative invasive options to the EVD.

C. Limitations

We acknowledge several limitations of this research. First, the sample size of this study was small, consisting of only 10 patients, which may impede our ability to generalize model predictions to a broader population. Second, our application of a selective pre-processing scheme to discern high-quality waveform data could introduce subjectivity and bias into the training data, potentially compromising model performance and robustness. Finally, our study lacked external validation through the use of an independent dataset, thereby restricting our ability to evaluate the generalizability and reliability of the findings beyond the data sourced from the MIMIC III Waveform Database.

D. Future Direction

Forthcoming endeavors will concentrate on augmenting both the sample size of evaluated patients, and further examining differences in model performances. Although 10 patients were used in this study, their data quality proved invaluable in drawing clinically significant conclusions and correlations. Nevertheless, we plan to expand our inclusion criteria by accommodating missing values (presently, five-minute intervals that comprise any missing values) and adding additional patients, thereby enhancing patient representation and improving model generalizability. We will also test the models on an external dataset outside of the MIMIC III Waveform Database to assess the generalizability of model performance. In particular, we intend to externally validate our data on a more homogenous group of TBI patients using the CENTER-TBI database. Finally, we are continuing to fine tune our individual deep learning models and are examining methods for combining the strengths of each model into a 'super-ensemble' learner to produce more optimal results. In doing so, we seek to utilize the strengths of each model to improve patient generalizability.

E. Supplementary Section

Supplementary Tables 1A-1F display the performance of each of the six deep learning models in all ten individual patients within the single-patient analysis. Supplementary Table 2A and 2B contains complete significance testing information when comparing these models against. Finally, Supplementary Figure 1 displays predicted versus ground truth predictions for each model in the multi-patient analysis. All tables and figures can be found after the references.

VII. References

[1] C. A. Taylor, J. M. Bell, M. J. Breiding and L. Xu, "Traumatic Brain Injury-Related Emergency Department Visits, Hospitalizations, and Deaths - United States, 2007 and 2013," MMWR Surveill Summ, vol. 66, no. 9, pp. 1-16, Mar. 2017.

[2] Centers for Disease Control and Prevention, "National Center for Health Statistics: Mortality Data on CDC WONDER," accessed April 2023, https://wonder.cdc.gov/mcd.html.

[3] L. Shen, Z. Wang, Z. Su, S. Qiu, J. Xu, Y. Zhou, A. Yan, R. Yin, B. Lu, X. Nie, S. Zhao, and R. Yan, "Effects of Intracranial Pressure Monitoring on Mortality in Patients with Severe Traumatic Brain Injury: A Meta-Analysis," PLoS One, vol. 11, no. 12, pp. e0168901, Dec. 2016, https://doi.org/10.1371/journal.pone.0168901.

[4] V. L. Pinto, P. Tadi, and A. Adeyinka, "Increased Intracranial Pressure," StatPearls, updated Aug. 4, 2021. Available: https://www.ncbi.nlm.nih.gov/books/NBK482119/

[5] M. J. Albeck, S. E. Borgesen, F. Gjerris, et al., "Intracranial pressure and cerebrospinal fluid outflow conductance in healthy subjects," J. Neurosurg., vol. 74, pp. 597-600, Apr. 1991.

[6] L. Rangel-Castilla, S. Gopinath, and C. S. Robertson, "Management of intracranial hypertension," Neurologic Clinics, vol. 26, no. 2, pp. 521-x, May 2008.

[7] M. H. Wilson, "Monro-Kellie 2.0: The dynamic vascular and venous pathophysiological components of intracranial pressure," Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism, vol. 36, no. 8, pp. 1338-1350, Aug. 2016. doi: 10.1177/0271678X16648711.

[8] X. Zhang, J. E. Medow, B. J. Iskandar, F. Wang, M. Shokoueinejad, J. Koueik and J. G. Webster, "Invasive and noninvasive means of measuring intracranial pressure: a review," Physiological measurement, vol. 38, no. 8, pp. R143-R182, Aug. 2017, doi: 10.1088/1361-6579/aa7256.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[9] M. Dey, J. Jaffe, A. Stadnik, and I. A. Awad, "External ventricular drainage for intraventricular hemorrhage," Current neurology and neuroscience reports, vol. 12, no. 1, pp. 24–33, 2012.

[10] A. Silverman and N. H. Petersen, "Physiology, Cerebral Autoregulation," StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, updated Mar 15, 2023. Available: https://www.ncbi.nlm.nih.gov/books/NBK553183/

[11] J. T. Levis, "ECG Diagnosis: Deep T Wave Inversions Associated with Intracranial Hemorrhage," The Permanente journal, vol. 21, pp. 16-049, 2017. doi: 10.7812/TPP/16- 049.

[12] H. Liu, F. Pan, X. Lei, J. Hui, R. Gong, J. Feng, and D. Zheng, "Effect of intracranial pressure on photoplethysmographic waveform in different cerebral perfusion territories: A computational study," Frontiers in Physiology, vol. 14, article no. 1085871, 2023, doi: 10.3389/fphys.2023.1085871.

[13] F. M. Kashif, G. C. Verghese, V. Novak, M. Czosnyka, and T. Heldt, "Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure," Sci. Transl. Med., vol. 4, no. 129, p. 129ra44, Apr. 2012, doi: 10.1126/scitranslmed.3003249.

[14] K. B. Evensen, M. O'Rourke, F. Prieur, et al., "Noninvasive Estimation of the Intracranial Pressure Waveform from the Central Arterial Blood Pressure Waveform in Idiopathic Normal Pressure Hydrocephalus Patients," Sci. Rep., vol. 8, no. 1, p. 4714, Mar. 2018, doi: 10.1038/s41598-018-23142-7.

[15] P. Wijayatunga, L. D. Koskinen, and N. Sundström, "Probabilistic prediction of increased intracranial pressure in patients with severe traumatic brain injury," Scientific Reports, vol. 12, no. 1, p. 9600, 2022. https://doi.org/10.1038/s41598-022-13732-x

[16] H. -J. Lee, H. Kim, Y. -T. Kim, K. Won, M. Czosnyka and D. -J. Kim, "Prediction of Life-Threatening Intracranial Hypertension During the Acute Phase of Traumatic Brain Injury Using Machine Learning," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 10, pp. 3967-3976, Oct. 2021, doi: 10.1109/JBHI.2021.3085881

[17] A. Uryga, A. Ziółkowski, A. Kazimierska, A. Pudełko, C. Mataczyński, E. W. Lang, M. Czosnyka, M. Kasprowicz, CENTER-TBI High-Resolution ICU (HR ICU) Sub-Study Participants and Investigators, and CENTER-TBI High-Resolution Sub-Study Participants and Investigators, "Analysis of intracranial pressure pulse waveform in traumatic brain injury patients: a CENTER-TBI study," J. Neurosurg., vol. Online, pp. 1- 11, Oct. 2022, Art no. 2022.10.JNS221523, doi: 10.3171/2022.10.JNS221523.

[18] D. Petrov, S. P. Miranda, R. Balu, C. Wathen, A. Vaz, V. Mohan, C. Colon, and R. Diaz- Arrastia, "Prediction of intracranial pressure crises after severe traumatic brain injury using machine learning algorithms," J. Neurosurg., vol. Online, pp. 1-8, Dec. 2022, Art no. 2022.12.JNS221860, doi: 10.3171/2022.12.JNS221860.

[19] G. Ye, V. Balasubramanian, J. K. Li, and M. Kaya, "Machine Learning-Based Continuous Intracranial Pressure Prediction for Traumatic Injury Patients," IEEE J. Transl. Eng. Health Med., vol. 10, p. 4901008, 2022, doi: 10.1109/JTEHM.2022.3179874.

[20] T. Goto, K. Furihata, and K. Hongo, "Natural resonance frequency of the brain depends on only intracranial pressure: clinical research," Sci. Rep., vol. 10, no. 1, p. 2526, Feb. 2020, doi: 10.1038/s41598-020-59376-7.

[21] A. E. Johnson, T. J. Pollard, L. Shen, L. W. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Sci. Data, vol. 3, p. 160035, May 2016, doi: 10.1038/sdata.2016.35.

[22] S. Meek and F. Morris, "ABC of clinical electrocardiography. Introduction. I-Leads, rate, rhythm, and cardiac axis," BMJ (Clin. Res. Ed.), vol. 324, no. 7334, pp. 415-418, Feb. 2002, doi: 10.1136/bmj.324.7334.415.

[23] R. D. Stevens, M. Shoykhet, and R. Cadena, "Emergency Neurological Life Support: Intracranial Hypertension and Herniation," Neurocrit. Care, vol. 23 Suppl 2, no. Suppl 2, pp. S76-S82, Dec. 2015, doi: 10.1007/s12028-015-0168-z.

[24] Dynamic Time Warping Information Retrieval for Music and Motion, Berlin, Heidelberg, Springer, 2007, pp. 69-84. doi: 10.1007/978-3-540-74048-3_4.

[25] A. Nayebi, S. Tipirneni, B. Foreman, J. Ratcliff, C. K. Reddy, and V. Subbian, "Recurrent Neural Network based Time-Series Modeling for Long-term Prognosis Following Acute Traumatic Brain Injury," in AMIA Symposium, 2021, pp. 900-909.

[26] D. Alis, C. Alis, M. Yergin, et al., "A joint convolutional-recurrent neural network with an attention mechanism for detecting intracranial hemorrhage on noncontrast head CT," Sci. Rep., vol. 12, no. 1, p. 2084, 2022, doi: 10.1038/s41598-022-05872-x.

[27] S. Wang, C. Shao, J. Zhang, et al., "Traffic flow prediction using bi-directional gated recurrent unit method," Urban Info, vol. 1, no. 1, p. 16, 2022, doi: 10.1007/s44212-022- 00015-z.

[28] C. Li and G. Qian, "Stock Price Prediction Using a Frequency Decomposition Based GRU Transformer Neural Network," Appl. Sci., vol. 13, no. 1, p. 222, 2023, doi: 10.3390/app13010222.

[29] S. Gao, J. Yang, T. Shen, and W. Jiang, "A Parallel Feature Fusion Network Combining GRU and CNN for Motor Imagery EEG Decoding," Brain Sci., vol. 12, no. 9, p. 1233, 2022, doi: 10.3390/brainsci12091233.

[30] S. Rudregowda, S. Tippannavar, N. Sharma, M. Rashid, K. Bhat, and A. Rana, "EEG- based

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Brain Wave Recognition using GRU and LSTM," in 2022 International Conference on Innovative Computing and Communication (ICICC), 2022, pp. 1-5, doi: 10.1109/IC3I56241.2022.10072419.

[31] A. Paul, "Prediction of missing EEG channel waveform using LSTM," in 2020 4th International Conference on Computational Intelligence and Networks (CINE), Kolkata, India, 2020, pp. 1-6, doi: 10.1109/CINE48825.2020.234393.

[32] X. Ma, S. Qiu, Y. Zhang, X. Lian, and H. He, "Predicting Epileptic Seizures from Intracranial EEG Using LSTM-Based Multi-task Learning," in Pattern Recognition and Computer Vision, PRCV 2018, Lecture Notes in Computer Science, vol. 11257, Springer, Cham, 2018, doi: 10.1007/978-3-030-03335-4$_1$4.

[33] W. Liu, S. Du, N. Pang, L. Zhang, G. Sun, H. Xiao, Q. Zhao, L. Xu, Y. Yao, J. Alastruey, and A. Avolio, "Central Aortic Blood Pressure Waveform Estimation Based on Temporal Convolutional Network," in IEEE Journal of Biomedical and Health Informatics, vol. PP, 2023, doi: 10.1109/JBHI.2023.3268886.

[34] K. Zhou, Z. Yin, Y. Peng, and Z. Zeng, "Methods for Continuous Blood Pressure Estimation Using Temporal Convolutional Neural Networks and Ensemble Empirical Mode Decomposition," in Electronics, vol. 11, no. 9, p. 1378, 2022, doi: 10.3390/electronics11091378.

[35] Y. K. Musallam, N. I. AlFassam, G. Muhammad, S. U. Amin, M. Alsulaiman, W. Abdul, H. Altaheri, M. A. Bencherif, and M. Algabri, "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," Biomedical Signal Processing and Control, vol. 69, p. 102826, 2021, doi: 10.1016/j.bspc.2021.102826.

[36] A. Salami, J. Andreu-Perez, and H. Gillmeister, "EEG-ITNet: An Explainable Inception Temporal Convolutional Network for Motor Imagery Classification," in IEEE Access, vol. 10, pp. 36672-36685, 2022. doi: 10.1109/ACCESS.2022.3161489.

[37] B.L. Hill, N. Rakocz, Á. Rudas, et al., "Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning," Scientific Reports, vol. 11, no. 1, p. 15755, 2021, doi: 10.1038/s41598-021-94913-y.

[38] S. M. Mousavi, W. L. Ellsworth, W. Zhu, Y. Shi, A. M. Zarei and B. A. Romanowicz, "Earthquake Transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking," in Nature Communications, vol. 11, no. 1, p. 3952, 2020, doi: 10.1038/s41467-020-17591-w.

[39] P. Verma and C. Chafe, "A Generative Model for Raw Audio Using Transformer Architectures," in 2021 24th International Conference on Digital Audio Effects (DAFx), Vienna, Austria, 2021, pp. 230-237, doi: 10.23919/DAFx51585.2021.9768298.

[40] R. A. Saleh and A. K. M. E. Saleh, "Statistical Properties of the log-cosh Loss Function Used in Machine Learning," arXiv preprint arXiv:2208.04564, 2022. [Online]. Available: https://arxiv.org/abs/2208.04564.

[41] P. Chen, G. Chen and S. Zhang, "Log Hyperbolic Cosine Loss Improves Variational Auto-Encoder," 2019. [Online]. Available: https://openreview.net/forum?id=rkglvsC9Ym.

[42] X. Xu, J. Li, Y. Yang, and F. Shen, "Toward Effective Intrusion Detection Using Log- Cosh Conditional Variational Autoencoder," in IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6187-6196, Apr. 2021. doi: 10.1109/JIOT.2020.3034621.

[43] S. Gopali, F. Abri, S. Siami-Namini and A. S. Namin, "A Comparison of TCN and LSTM Models in Detecting Anomalies in Time Series Data," in 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 2415-2420, doi: 10.1109/BigData52589.2021.9671488.

[44] D. Cardim, C. Robba, M. Bohdanowicz, J. Donnelly, B. Cabella, X. Liu, M. Cabeleira, P. Smielewski, B. Schmidt, and M. Czosnyka, "Non-invasive Monitoring of Intracranial Pressure Using Transcranial Doppler Ultrasonography: Is It Possible?," Neurocrit Care, vol. 25, no. 3, pp. 473-491, Dec. 2016, doi: 10.1007/s12028-016-0258-6.

[45] M. Megjhani, K. Terilli, B. Weinerman, D. Nametz, S. B. Kwon, A. Velazquez, S. Ghoshal, D. J. Roh, S. Agarwal, E. S. Connolly Jr, J. Claassen, and S. Park, "A Deep Learning Framework for Deriving Non-Invasive Intracranial Pressure Waveforms from Transcranial Doppler," Annals of Neurology, vol. 10.1002/ana.26682, 2023, doi: 10.1002/ana.26682. [Online]. Available: https://doi.org/10.1002/ana.26682

[46] P. H. Raboel, J. Bartek Jr, M. Andresen, B. M. Bellander and B. Romner, "Intracranial Pressure Monitoring: Invasive versus Non-Invasive Methods-A Review," Crit. Care Res. Pract., vol. 2012, pp. 950393, 2012, doi: 10.1155/2012/950393.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Supplemental Table 1A. Single Patient Analysis Results for Simple RNN

	Pat 1	**Pat 2**	**Pat 3**	**Pat 4**	**Pat 5**	**Pat 6**	**Pat 7**	**Pat 8**	**Pat 9**	**Pat 10**	**Mean ± Std**	**Median**
Logcosh Test Loss (Train Loss) in mmHg	3.47 (3.58)	2.49 (2.47)	1.87 (1.85)	2.28 (2.37)	2.52 (2.56)	2.36 (2.61)	3.24 (3.29)	3.05 (3.09)	1.35 (1.40)	3.11 (3.63)	2.57 ±0.66 (2.69 ± 0.73)	2.51 (2.59)
MAE Test Loss (Train Loss) in mmHg	4.10 (4.25)	3.09 (3.18)	2.46 (2.53)	2.87 (2.94)	3.11 (3.16)	2.94 (3.25)	3.86 (4.04)	3.67 (3.83)	1.89 (1.96)	3.74 (4.41)	3.17 ±0.68 (3.36 ± 0.78)	3.10 (3.22)
RMSE Test Loss (Train Loss) in mmHg	5.17 (5.28)	4.08 (4.29)	3.07 (3.16)	4.38 (4.64)	4.27 (4.35)	4.16 (4.60)	4.92 (5.16)	4.74 (4.96)	2.74 (2.85)	5.04 (5.94)	4.26 ±0.81 (4.52 ± 0.94)	4.33 (4.62)

Supplemental Table 1B. Single Patient Analysis Results for GRU

	Pat 1	**Pat 2**	**Pat 3**	**Pat 4**	**Pat 5**	**Pat 6**	**Pat 7**	**Pat 8**	**Pat 9**	**Pat 10**	**Mean ± Std**	**Median**
Logcosh Test Loss (Train Loss) in mmHg	2.05 (1.81)	0.92 (0.83)	0.32 (0.24)	1.06 (0.87)	0.83 (0.53)	0.86 (0.89)	0.53 (0.57)	0.40 (0.34)	0.40 (0.38)	1.24 (0.84)	0.86 ±0.52 (0.73 ± 0.45)	0.85 (0.70)
MAE Test Loss (Train Loss) in mmHg	2.63 (2.32)	1.42 (1.28)	0.67 (0.47)	1.54 (1.40)	1.32 (0.83)	1.30 (1.39)	0.94 (0.96)	0.78 (0.73)	0.79 (0.72)	1.75 (1.17)	1.31 ±0.59 (1.13 ± 0.52)	1.31 (1.07)
RMSE Test Loss (Train Loss) in mmHg	3.48 (3.10)	2.03 (1.84)	1.07 (0.80)	2.59 (2.37)	1.89 (1.19)	2.03 (2.16)	1.32 (1.44)	1.26 (1.13)	1.23 (1.11)	2.59 (1.74)	1.95 ±0.77 (1.69 ± 0.70)	1.96 (1.59)

Supplemental Table 1C. Single Patient Analysis Results for LSTM

	Pat 1	**Pat 2**	**Pat 3**	**Pat 4**	**Pat 5**	**Pat 6**	**Pat 7**	**Pat 8**	**Pat 9**	**Pat 10**	**Mean ± Std**	**Median**
Logcosh Test Loss (Train Loss) in mmHg	2.02 (1.74)	0.90 (0.54)	0.48 (0.43)	1.00 (0.91)	0.83 (0.52)	0.89 (0.75)	0.47 (0.39)	0.83 (0.78)	1.62 (1.33)	1.80 (1.73)	1.08 ±0.54 (0.91 ± 0.51)	0.90 (0.77)
MAE Test Loss (Train Loss) in mmHg	2.61 (2.27)	1.38 (0.82)	0.88 (0.81)	1.47 (1.41)	1.32 (0.81)	1.34 (1.17)	0.88 (0.67)	1.91 (1.71)	2.15 (1.79)	2.36 (2.29)	1.63 ±0.60 (1.38 ± 0.62)	2.04 (1.50)
RMSE Test Loss (Train Loss) in mmHg	3.45 (2.98)	2.02 (1.21)	1.29 (1.27)	2.57 (2.47)	1.90 (1.16)	2.05 (1.73)	1.24 (0.97)	1.31 (1.21)	3.34 (2.43)	3.39 (3.23)	2.26 ±0.89 (1.87 ± 0.84)	2.04 (1.50)

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Supplemental Table 1D. Single Patient Analysis Results for TCN

	Pat 1	**Pat 2**	**Pat 3**	**Pat 4**	**Pat 5**	**Pat 6**	**Pat 7**	**Pat 8**	**Pat 9**	**Pat 10**	**Mean ± Std**	**Median**
Logcosh Test Loss (Train Loss) in mmHg	1.92 (0.96)	0.87 (0.48)	0.24 (0.08)	0.80 (0.34)	0.70 (0.44)	0.76 (0.47)	0.39 (0.29)	0.35 (0.16)	0.36 (0.23)	1.12 (0.65)	0.75 ±0.50 (0.41 ± 0.26)	0.73 (0.39)
MAE Test Loss (Train Loss) in mmHg	2.49 (1.21)	1.36 (0.71)	0.54 (0.19)	1.20 (0.51)	1.16 (0.73)	1.20 (0.76)	0.77 (0.55)	0.72 (0.34)	0.74 (0.47)	1.63 (0.95)	1.18 ±0.57 (0.64 ± 0.30)	1.18 (0.63)
RMSE Test Loss (Train Loss) in mmHg	3.41 (1.75)	2.01 (1.08)	0.94 (0.38)	2.34 (1.08)	1.71 (1.13)	1.92 (1.26)	1.14 (0.84)	1.16 (0.54)	1.14 (0.74)	2.49 (1.44)	1.83 ±0.78 (1.02 ± 0.41)	1.82 (1.08)

Supplemental Table 1E. Single Patient Analysis Results for V-NET

	Pat 1	**Pat 2**	**Pat 3**	**Pat 4**	**Pat 5**	**Pat 6**	**Pat 7**	**Pat 8**	**Pat 9**	**Pat 10**	**Mean ± Std**	**Median**
Logcosh Test Loss (Train Loss) in mmHg	1.94 (1.16)	0.88 (0.48)	0.25 (0.05)	0.90 (0.47)	0.75 (0.43)	0.75 (0.34)	0.41 (0.17)	0.37 (0.13)	0.37 (0.16)	1.17 (0.48)	0.80 ±0.50 (0.39 ± 0.32)	0.75 (0.39)
MAE Test Loss (Train Loss) in mmHg	2.52 (1.49)	1.36 (0.69)	0.57 (0.11)	1.34 (0.75)	1.23 (0.68)	1.18 (0.53)	0.80 (0.28)	0.75 (0.26)	0.76 (0.37)	1.68 (0.67)	1.22 ±0.57 (0.58 ± 0.39)	1.21 (0.61)
RMSE Test Loss (Train Loss) in mmHg	3.38 (1.94)	2.00 (1.05)	0.98 (0.19)	2.45 (1.37)	1.78 (0.95)	1.90 (0.86)	1.19 (0.42)	1.21 (0.44)	1.15 (0.58)	2.54 (1.07)	1.86 ±0.77 (0.89 ± 0.52)	1.84 (0.91)

Supplemental Table 1F. Single Patient Analysis Results for Transformer

	Pat 1	**Pat 2**	**Pat 3**	**Pat 4**	**Pat 5**	**Pat 6**	**Pat 7**	**Pat 8**	**Pat 9**	**Pat 10**	**Mean ± Std**	**Median**
Logcosh Test Loss (Train Loss) in mmHg	3.21 (3.30)	1.91 (2.03)	0.92 (0.94)	2.41 (2.56)	1.57 (1.66)	3.21 (3.25)	2.07 (2.11)	1.65 (1.83)	1.19 (1.19)	2.63 (2.62)	2.08 ±0.79 (2.15 ± 0.79)	1.99 (2.07)
MAE Test Loss (Train Loss) in mmHg	3.84 (3.95)	2.50 (2.66)	1.43 (1.45)	2.99 (3.23)	2.13 (2.24)	3.82 (3.87)	2.65 (2.71)	2.21 (2.41)	1.71 (1.71)	3.22 (3.21)	2.65 ±0.82 (2.74 ± 0.84)	2.58 (2.69)
RMSE Test Loss (Train Loss) in mmHg	4.87 (5.01)	3.34 (3.56)	1.90 (1.83)	4.53 (4.89)	2.91 (3.08)	5.00 (5.07)	3.58 (3.66)	3.00 (3.27)	2.46 (2.46)	4.51 (4.50)	3.61 ±1.07 (3.73 ± 1.12)	3.46 (3.61)

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

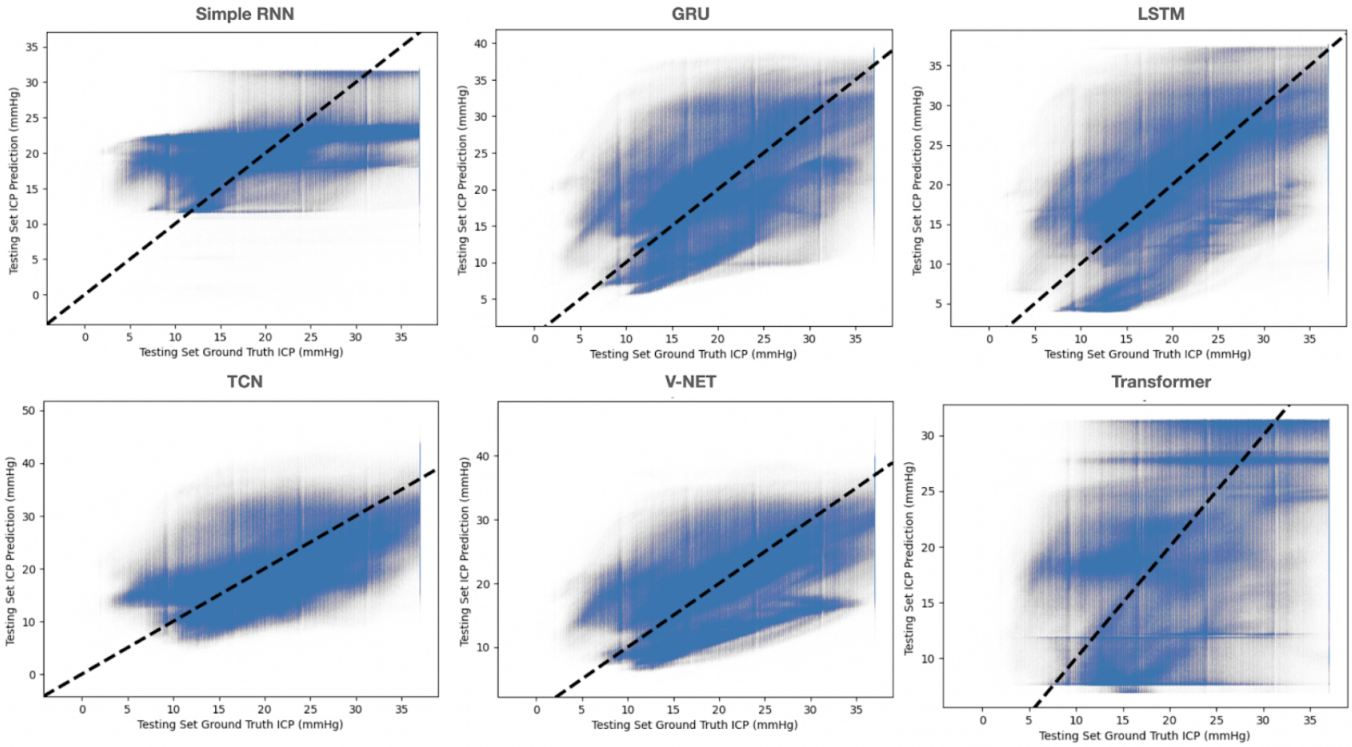
Supplemental Table 2A. Single Patient Paired One-Tailed T-Test Between Models (Model 1 MAE > Model 2 MAE)

y	Model 1	Model 2	n1	n2	statistic	df	p	p.adj	p.adj.signif
mae	GRU	LSTM	10	10	-1.88	9	9.53e-01	1.00e+00	ns
mae	GRU	RNN	10	10	-9.70	9	1.00e+00	1.00e+00	ns
mae	GRU	TCN	10	10	4.99	9	3.72e-04	6.00e-03	**
mae	GRU	Transformer	10	10	-8.11	9	1.00e+00	1.00e+00	ns
mae	GRU	VNET	10	10	5.79	9	1.00e+00	1.00e+00	**
mae	LSTM	RNN	10	10	-6.24	9	1.00e+00	1.00e+00	ns
mae	LSTM	TCN	10	10	2.88	9	9.00e-03	1.35e-01	ns
mae	LSTM	Transformer	10	10	-3.97	9	9.98e-01	1.00e+00	ns
mae	LSTM	VNET	10	10	2.62	9	1.40e-02	2.08e-01	ns
mae	RNN	TCN	10	10	10.51	9	1.18e-06	1.77e-05	****
mae	RNN	Transformer	10	10	2.37	9	2.10e-02	3.15e-01	ns
mae	RNN	VNET	10	10	10.26	9	1.44e-06	2.16e-05	****
mae	TCN	Transformer	10	10	-8.70	9	1.00e+00	1.00e+00	ns
mae	TCN	VNET	10	10	-2.77	9	9.89e-01	1.00e+00	ns
mae	Transformer	VNET	10	10	8.36	9	7.74e-06	1.16e-04	***

Supplemental Table 2B. Single Patient Paired One-Tailed T-Test Between Models (Model 1 MAE < Model 2 MAE)

y	Model 1	Model 2	n1	n2	statistic	df	p	p.adj	p.adj.signif
mae	GRU	LSTM	10	10	-1.88	9	4.70e-02	7.02e-01	ns
mae	GRU	RNN	10	10	-9.70	9	2.31e-06	3.46e-05	****
mae	GRU	TCN	10	10	4.99	9	1.00e+00	1.00e+00	ns
mae	GRU	Transformer	10	10	-8.11	9	9.96e-06	1.49e-04	***
mae	GRU	VNET	10	10	5.79	9	1.00e+00	1.00e+00	ns
mae	LSTM	RNN	10	10	-6.24	9	7.59e-05	1.00e-03	**
mae	LSTM	TCN	10	10	2.88	9	9.91e-01	1.00e+00	ns
mae	LSTM	Transformer	10	10	-3.97	9	2.00e-03	2.40e-02	*
mae	LSTM	VNET	10	10	2.62	9	9.86e-01	1.00e+00	ns
mae	RNN	TCN	10	10	10.51	9	1.00e+00	1.00e+00	ns
mae	RNN	Transformer	10	10	2.37	9	9.79e-01	1.00e+00	ns
mae	RNN	VNET	10	10	10.26	9	1.00e+00	1.00e+00	ns
mae	TCN	Transformer	10	10	-8.70	9	5.64e-06	8.46e-05	****
mae	TCN	VNET	10	10	-2.77	9	1.10e-02	1.64e-01	ns
mae	Transformer	VNET	10	10	8.36	9	1.00e+00	1.00e+00	ns

medRxiv preprint doi: https://doi.org/10.1101/2023.05.16.23289747; this version posted July 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .



Supplemental Figure 1. Multi-Train Analysis Prediction vs Ground Truth for Each Model