Informative Data Science to Predict Pulmonary Symptomatology for Divers

Nicholas J. Napoli [1] , Szilard Beres [1] , David M. Fothergill [2] , John P. Florian [2]

[1] Dept of Electrical & Computer Engineering
Human Informatics and Predictive
Performance Optimization (HIPPO) Lab
University of Florida (Gainesville)

[2] Naval Submarine Medical Research Laboratory
Naval Submarine Base New London
Box 900 Groton, CT 06349

Background Photo AI Generated using ChatGPT

Disclaimer

- The **views** expressed in this presentation are **those of the authors** and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.
    - Appearance of a U.S. Government graphic **does not imply endorsement** of the agency which created the graphic.
- Drs. Fothergill and Florian are **employees of the U.S. Government**. This work was prepared as part of their official **duties**.
    - Title 17 U.S.C. §105 provides that **copyright** protection under this title is **not available** for any work of the U. S. Government. Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.
- Mention or display of any companies or commercial items in this presentation **does not imply government endorsement** of those companies or items. I have no conflicts of interest.
- The authors have no conflicts of interest to declare.

Background and Motivation

- Special Operation Forces (SOF) missions require extended duration single and repeated dives
- Repeated oxygen-rich dive exposures can produce pulmonary oxygen stress before overt toxicity
- No models exists to estimate a diver’s physiological readiness
- This study identifies personalized physiological variables that predict pulmonary symptoms associated with subsequent dives

U.S. National Archives #6484219

Photo AI Generated using ChatGPT

Pulmonary Symptoms: Inspiratory Burning, Chest Tightness, Dyspnea, Cough

Methods: Combining and Restructuring the Dataset

Unified Data Set

Dataset by Study Conditions

- Gas Condition
  - Oxygen vs Air
- Exercise Condition
  - Exercise vs Resting
- Pre vs Post Dive Physiology
- Immersion vs Air
- **Dive Day (1-5)**
- Pulmonary Symptoms
  - Cough
  - Dyspnea
  - Inspiratory Burning
  - Chest Tightness

1104 Dives x 140 Variables

Predicting Future Dive Day Data Set

Dive Number Mapping to Outcome

Symptoms on Dive N are predicted using physiological burden from Dive N−1 and physiological readiness immediately before Dive N.

Prior Dive Burden + Current Pre-Dive State → Current Dive Symptom Risk

General Physiological Variables

- Vascular/Endothelial
- Cardiac control
- Pulmonary
- Muscular
- Labs and Blood
- Demographics

This data structure addresses the research question "***Can a diver's pre and post physiological data determine if the diver will experience pulmonary symptomatology on their current dive?***"

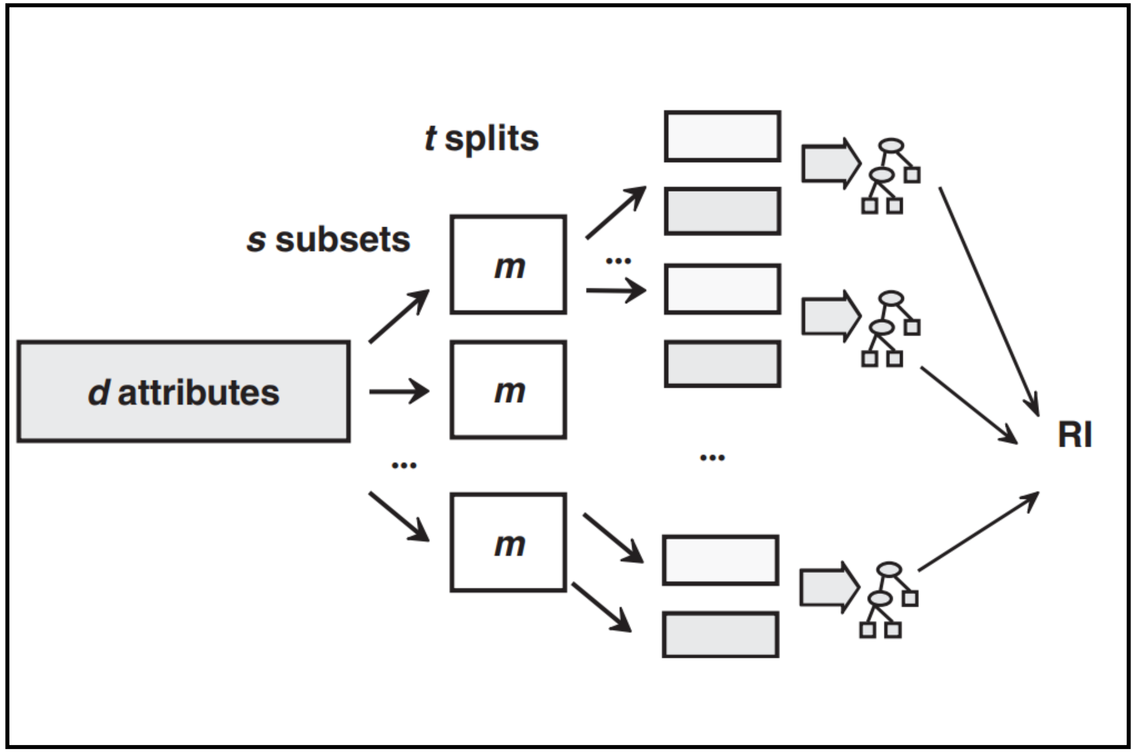
Methods: Analysis Across Entire Symptomatology Data

Monte Carlo Subspace Analysis

Constraints and Design:

- 3 Variables within a Model
- >40% of Available Variables Present
- A Min of 90 Samples per Model
- Class Imbalance Min at 20%
    - Adjusted with SMOTE
- Data Subsets of T-splits (e.g. 5000)
- Leave-One-Subject Cross Validation

Monte Carlo Subspace Analysis



Draminski M, Rada-Iglesias A, Enroth S, Wadelius C, Koronacki J, Komorowski J. **"Monte Carlo feature selection for supervised classification. Bioinformatics".** 2008 Jan 1;24(1)

SMOTE: Synthetic Minority Over-sampling Technique

Model Performance Metrics: F1-Score and Confusion Matrix

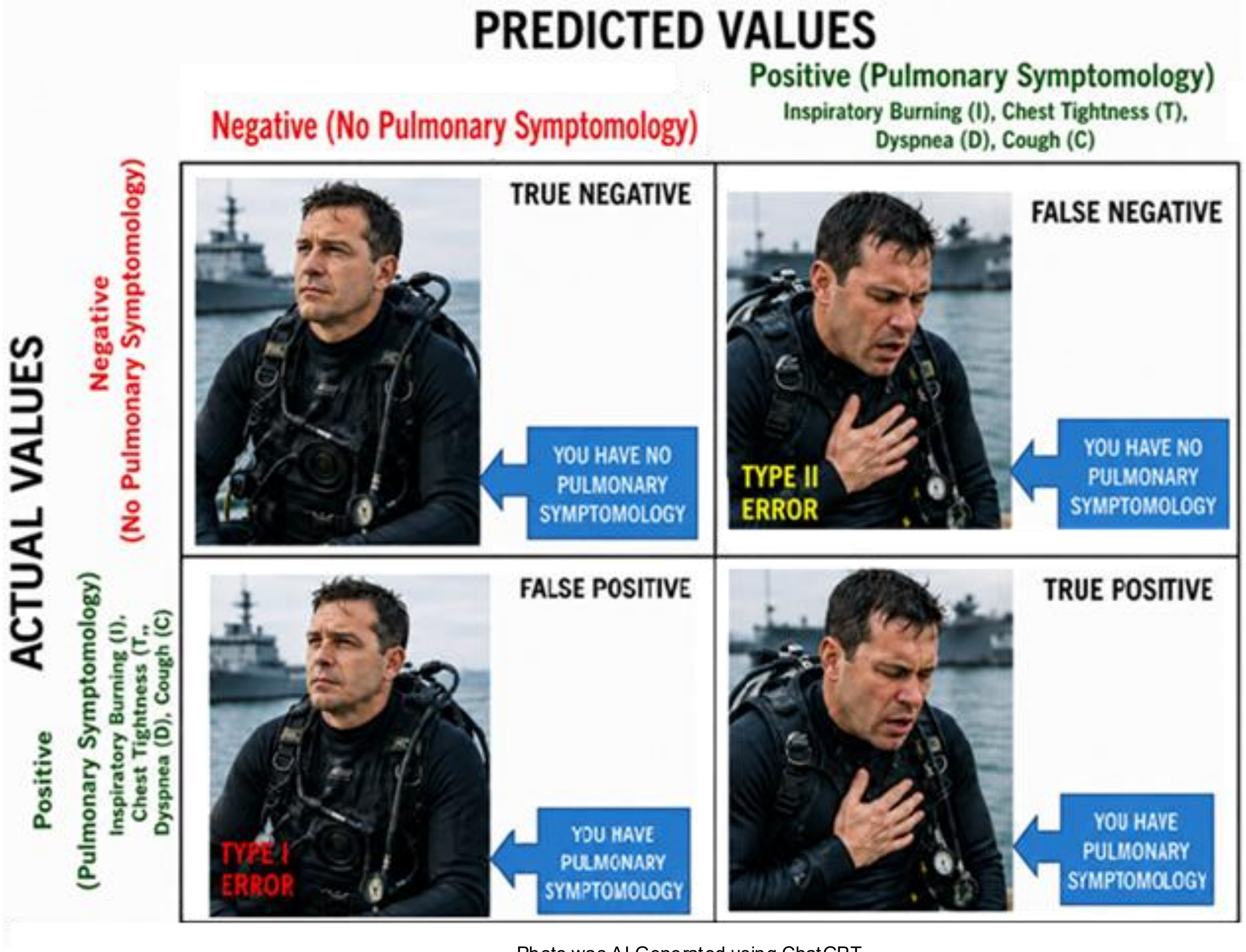


Photo was AI Generated using ChatGPT

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

We use F1 Score to evaluate model performance to focus on how well the model handles the minority class.

Predictive Modeling Roadmap

Military Diver Pulmonary Symptomatology

Modeling Approach A

Logistic Regression Predictive Models

Estimates the probability of pulmonary symptomology and quantifies each symptoms predictor

Use for: Risk scores, odds ratios, model coefficients

Modeling Approach B

Simple Tree Predictive Models

Converts model behavior into readable decision paths for operational environments

Use for: Thresholds, if/then rules, interpretable branching

Background Photo AI Generated using ChatGPT

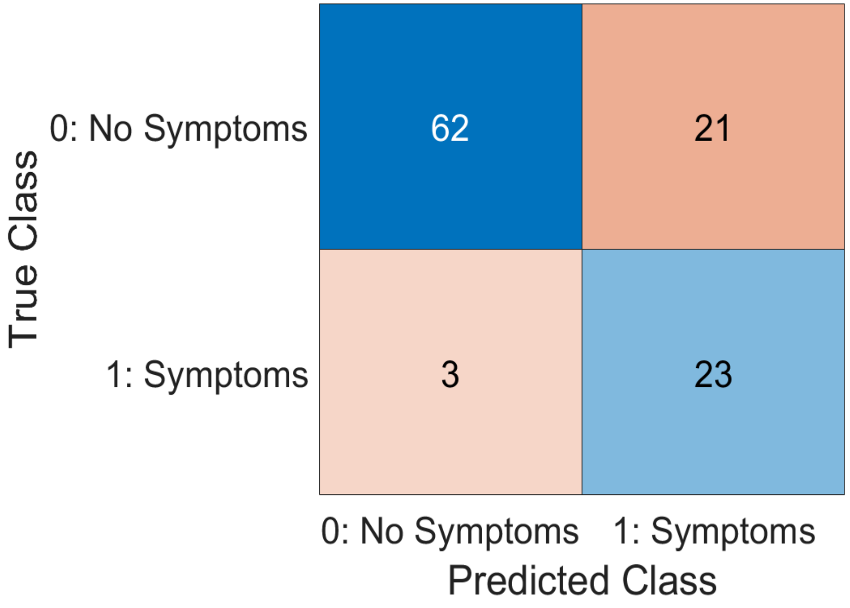
Models Predictive Performance and Stability

Logistic Regression Model for Symptomatology

Summary Distribution of Beta Coefficients

Variable	Mean	Median	SE	Max	Min
FEF25_75 Pre-Dive	0.718	0.689	0.0118	1.07	0.65
LF HRV Post	0.0047	0.0049	0.00016	1.0083	0.0012
NO_{Exp} Post	-0.423	-0.4114	0.0047	-0.397	-0.564

Confusion Matrix (F1-Score=0.74)



Acronyms

FEF 25_75: Forced Expired Flow (L/s)
LF HRV: Low Frequency Heart Rate Variability (ms^2)
NO_{Exp}: Expired Nitric Oxide (ppb)

The models exhibit consistent stability across-validation folds, underscoring the reliability and generalizability of the variables.

Interpreting Predictive Model for Pulmonary Symptomatology

Logistic Regression Model for Symptomatology

Understanding Risk: Each Coefficient Beta tells you how a predictor changes the log of the odds.

Force Expired Flow ($FEF_{25\text{-}75}$): (β1= 0.689, Odd Ratio=1.99)

- Variable Range: min: 1.55 L/s, mean: 3.86 L/s, max: 6.47 L/s
- A one-unit increase will increase the odds by 99%
- Suggests altered small-airway function may be linked to symptom development

LF Heart Rate Variability Post: (β2= 0.0049, Odd Ratio=1.005)

- Variable Range: min: 25.31 ms^2, mean: 144.41 ms^2, max: 363.41 ms^2
- Odds increase by 0.5% per unit, or approximately 63% per 100 units
- Indicates autonomic nervous system changes following diving exposure
- May reflect increased physiological stress or altered cardiovascular regulation

Expired Nitric Oxide Post Dive: (β3= -0.4114, Odd Ratio=0.66)

- Variable Range: min: 4 ppb, mean: 12.38 ppb, max: 53.00 ppb
- Odds decrease by 34% per unit, or approximately 87% over 5 units

Predictive Modeling Roadmap

Military Diver Pulmonary Symptomatology

Modeling Approach A

Logistic Regression Predictive Models

Estimates the probability of pulmonary symptomology and quantifies each symptoms predictor

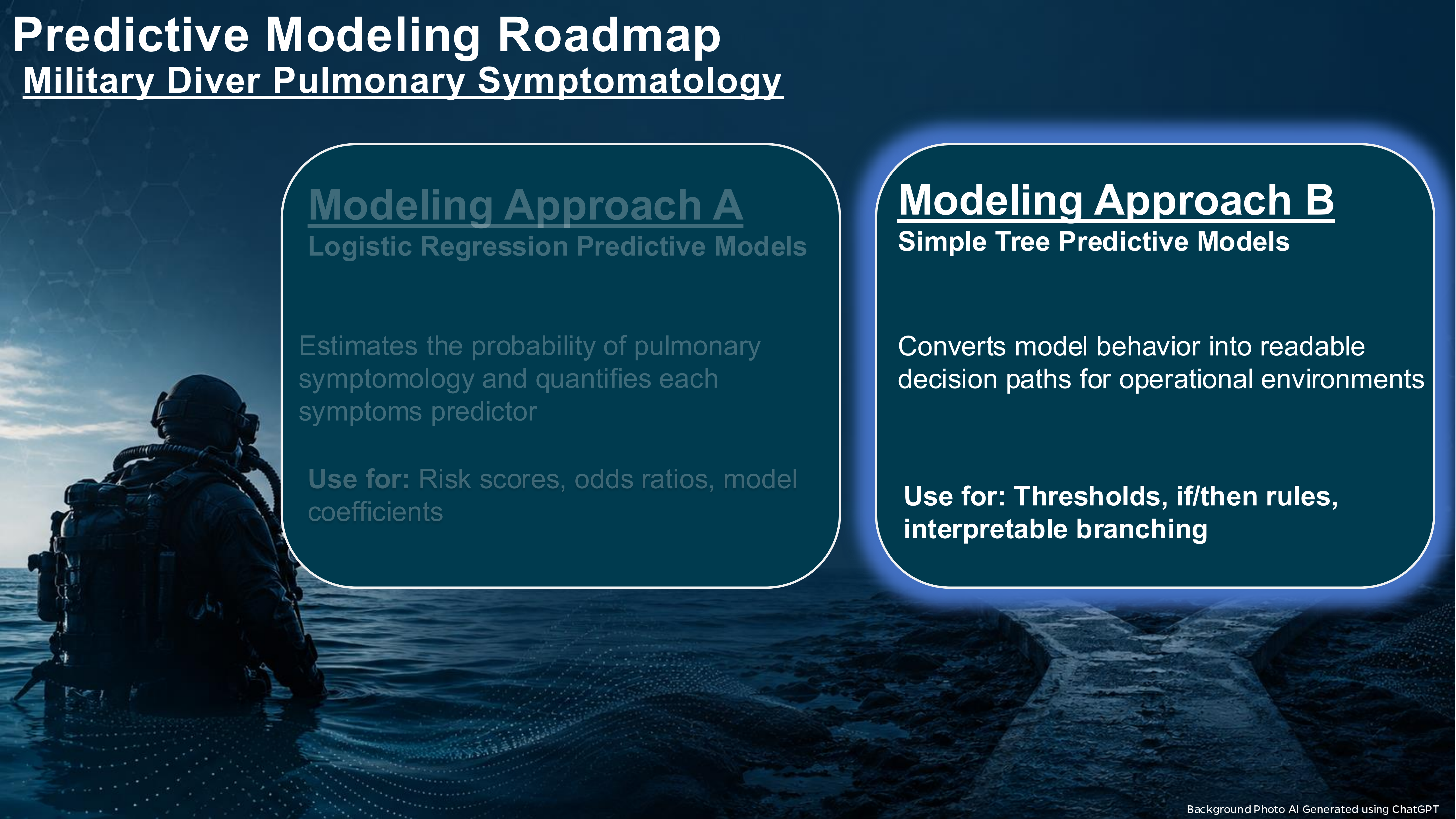
Use for: Risk scores, odds ratios, model coefficients

Modeling Approach B

Simple Tree Predictive Models

Converts model behavior into readable decision paths for operational environments

Use for: Thresholds, if/then rules, interpretable branching



Background Photo AI Generated using ChatGPT

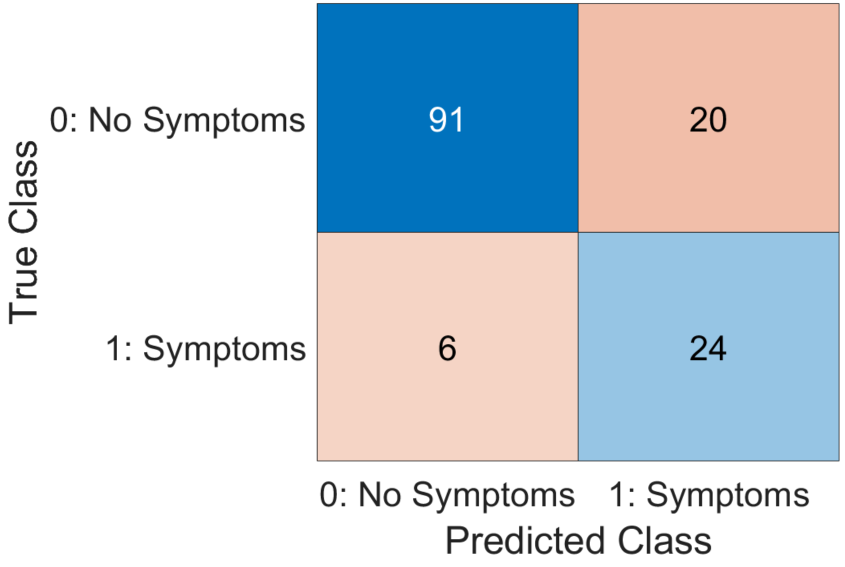
Model’s Predictive Performance and Stability

Simple Tree Model for Symptomatology

Summary Distribution of Decision Branches

Variable	Mean	Median	SE	Max	Min
NO_{Exp} (ppb)	14.99	14.94	0.04	15.94	14.8
USG Ratio	1.06	1.0067	0.0003	1.0083	1.0003
NO_{Exp} Ratio	1.752	1.603	0.065	2.602	1.168

Confusion Matrix
(F1-Score=0.747)



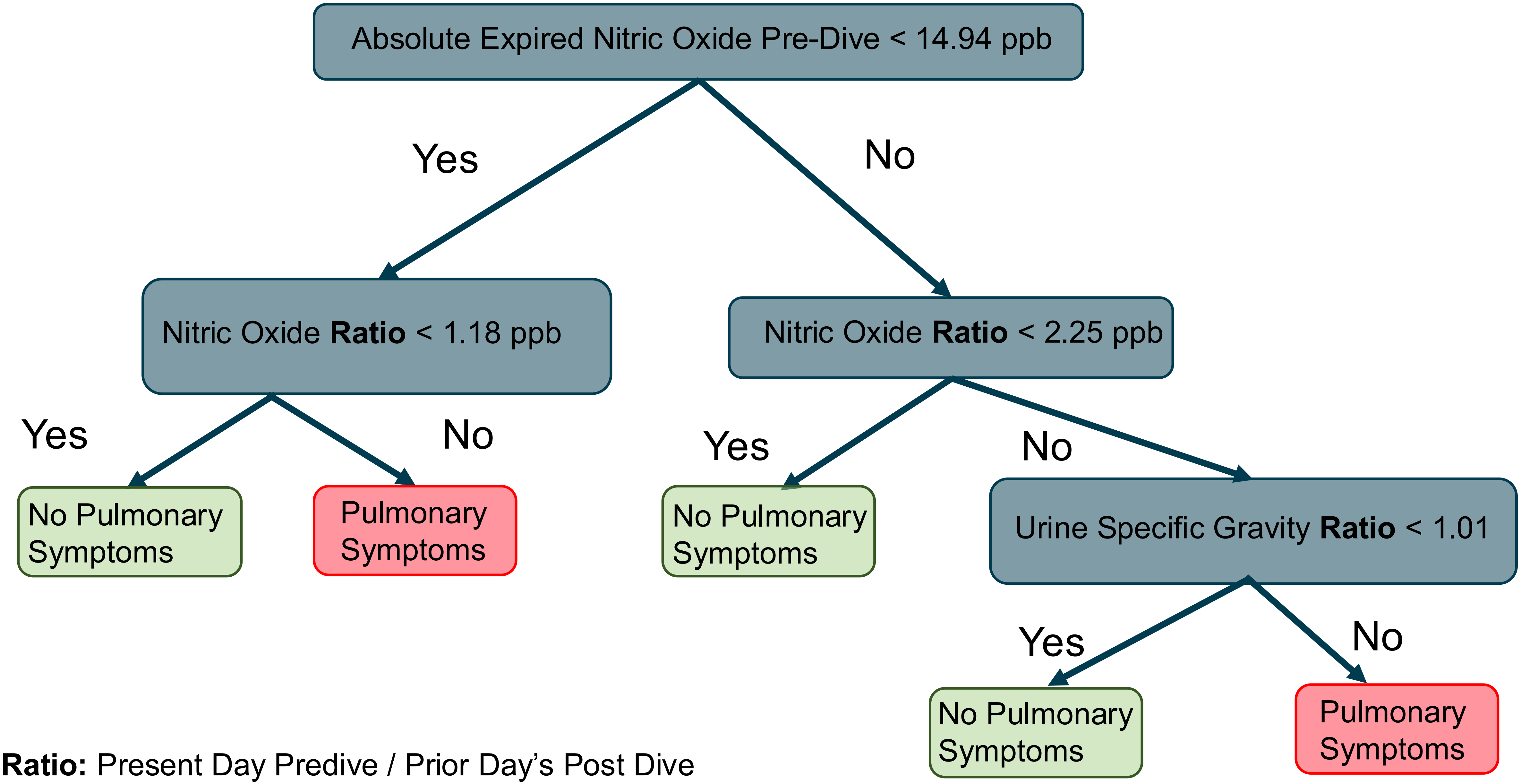
Acronyms

NO_{Exp}: Expired Nitric Oxide

USG Ratio: Urine Specific Gravity Ratio

The model exhibits stability across-validation folds, underscoring the reliability and generalizability of the predictive variables.

Interpreting Decision Tree for Pulmonary Symptomatology



Ratio: Present Day Predive / Prior Day's Post Dive

Physiological Interpretation

- Baseline NO_{Exp} may reflect airway physiological reserve to hyperbaric oxidative stress
- Divers with higher baseline NO_{Exp} tolerated larger post-dive NO_{Exp} reductions without developing symptoms*
- Symptom risk was associated with residual physiological burden from previous dive exposures
- Hydration and autonomic function may be linked to developing pulmonary symptoms

Bottom Line: Findings support a potential role of **airway inflammation and physiological stress** in symptom development following diving exposure.

.

*Fothergill, D.M., and Gertner, J.W. (2023). Exhaled Nitric Oxide and Pulmonary Oxygen Toxicity Susceptibility. Metabolites 13(8): 930.

Conclusion

At the onset of a dive, the risk of developing pulmonary symptoms increases when the following physiological factors are altered:

- Pulmonary oxidative stressors
  - NO_{Exp}
  - Forced Expired Flow (FEF_{25-75})
- Hydration state
  - Urine Specific Gravity Ratio
- Autonomic Response
  - LF Heart Rate Variability

Bottom Line: Characterizing the diver's physiological state entering the dive is pivotal for identifying pulmonary symptom risk.

Acknowledgements

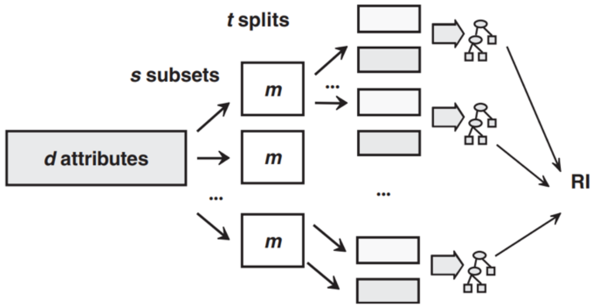
Office of Naval Research Summer Faculty Research & Sabbatical Leave Program

For more information:

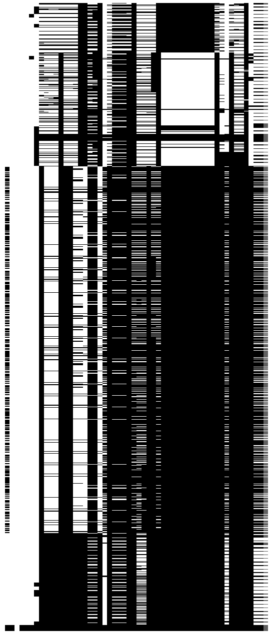
- n.napoli@ufl.edu

Appendix

Modeling Design Decisions for Outcome Variables



Data Base Structure And Constraints



Modeling Design Decisions for Outcome Variables

Design Decisions: Modeling on an Outcome Variable

1) Model Utility

Model Design Needs to be Translational

- Go-No-Go (Binary Response)
- Interpretable Model
  - Physiological Meaning
  - Inform Future Studies
  - Inform Mission Readiness

2) Design Constraints and Considerations

- Number of Samples
- Class Imbalances
- Coverage for Stability and Agnostic Model
  - # Predictor of Variables Associate to the Outcome
  - Across Studies (RIP, RIPX, etc)
  - Within Studies (RIP, RIPX, etc)
- Variables Need to be Physiological Based

Outcome (Response) Variables of Interest

1. Hand Grip Strength
   - Utility: SpecWar Tasks
   - Samples: 387
   - Class Balance: Threshold Required
2. Hand Grip Time to Fatigue
   - Utility: SpecWar Tasks
   - Samples: 387
   - Class Balance: Threshold Required
3. 85 % VO_{Max} Endurance Trial
   - Samples: 97 (Other Data tied to SD)
   - Class Balance: Threshold Required
4. Symptomatology (Total Sym)
   - **Samples: 399**
   - Class Balance: Sym/NoSym=95/304 **31.25%**
5. **Pulmonary Function !!!!! For white paper…**

The most methodologically sound initial strategy for developing a predictive model is to favor symptomology.

Data Base Structure And Constraints

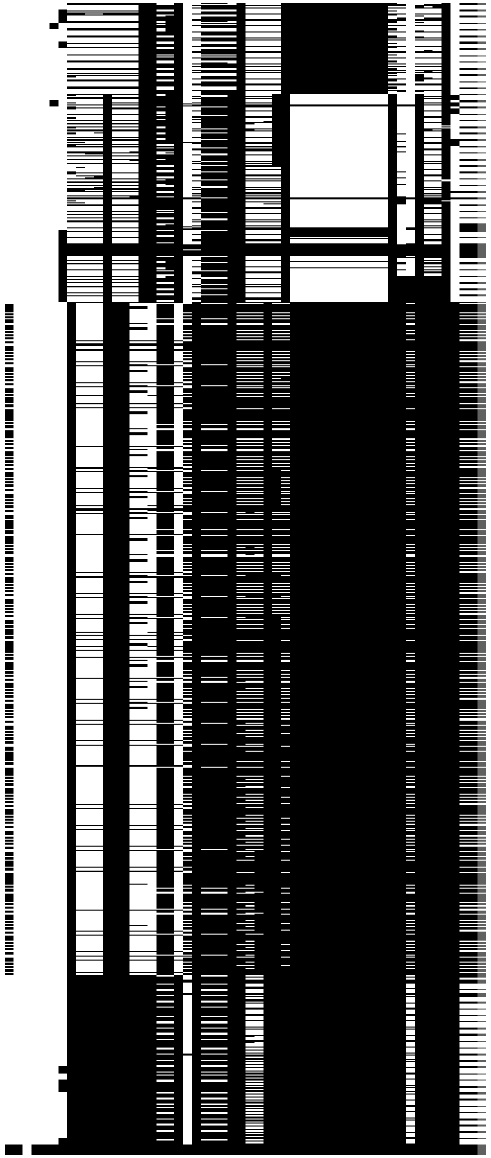
Overall Data and Different Data Sets

Total Dataset

- 2842 Pre and Post Dives Samples
- 140 total Variables
- Not Aligned Data
  - 65.81 % of the Data is Not Aligned ("Missing")
- Missing data has structure to the dataset

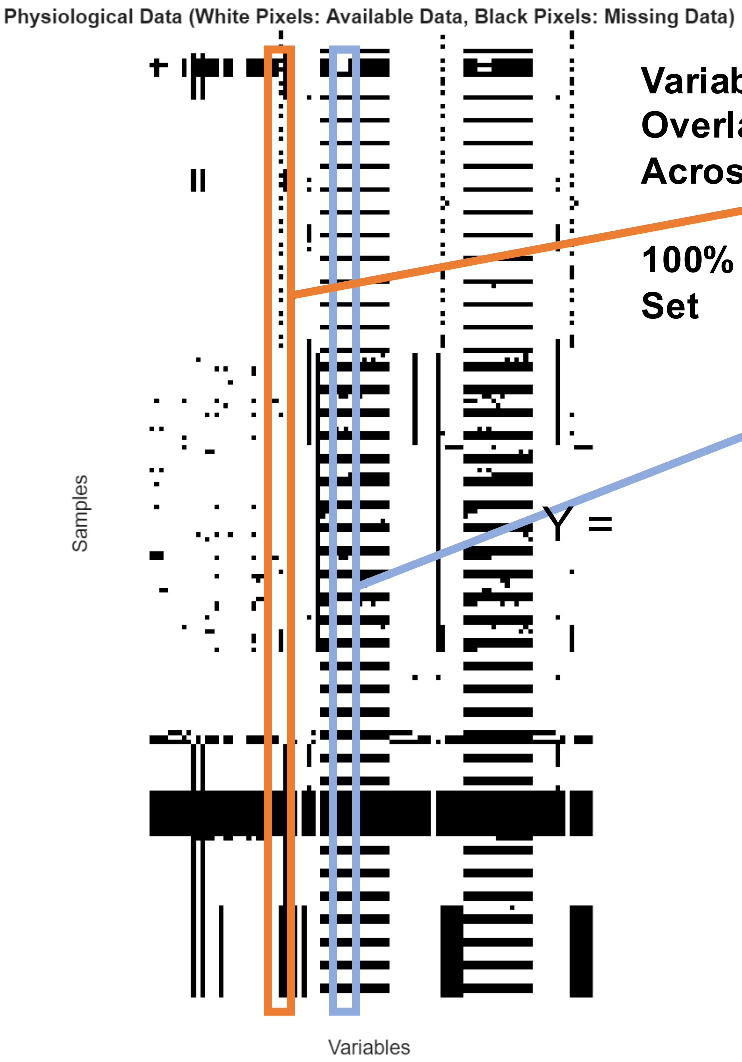
A Structure Data Set with Data Not Missing at Random

Black Pixels: Missing Data
White Pixels: Available Data



Bottomline: Although the dataset is large, there are limited amount of variables that align across studies which will impact the complexity of the modeling, the types of models, preprocessing of the data and research questions.

Non-Overlapping Variables Across Data Sets



Variable Overlaps Across Study

100% of Data Set

$$a_1 X_1 + a_2 X_2 = Y$$

Variable Inconsistent Overstudy Designs

70% of Data Set

Model is limited by the samples and is just the lowest amount of sampling

Sample #	Var 1	Var 2	Used Samples
1	21.05	0.60	1
2	21.05	0.64	2
3	21.05	0.76	3
4	21.05	0.64	4
5	NaN	0.60	
6	NaN	0.42	
7	NaN	0.45	
8	NaN	0.46	
9	NaN	1.03	
10	NaN	0.85	
11	NaN	0.63	
12	NaN	1.03	
13	16.53	0.70	5
14	16.53	1.00	6
15	16.53	1.00	7
16	16.53	0.67	8
17	23.39	0.44	9
18	23.39	0.62	10
19	23.39	1.00	11
20	23.39	0.67	12
21	13.56	0.54	13
22	13.56	0.64	14
23	13.56	0.44	15
24	13.56	0.55	16
25	0.66	NaN	
26	0.32	NaN	
27	NaN	NaN	
28	0.20	NaN	
29	16.24	0.55	17
30	16.24	0.89	18

While the dataset is sufficient for certain combinations, including too many variables can sharply reduce sample size and compromise model stability.

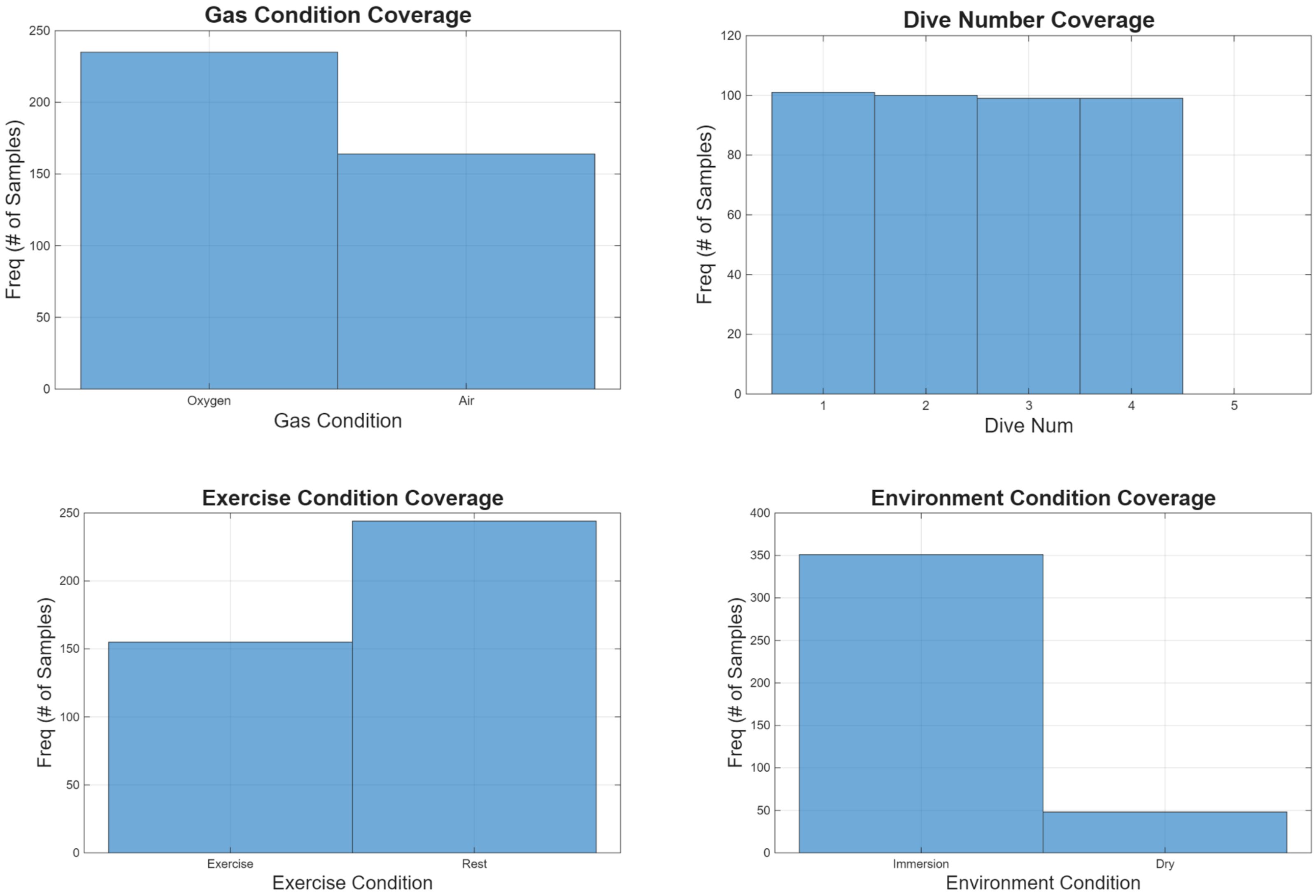
Reviewing Outcome Variable Coverage

Predicting Future Dive Day Data Set

Dataset by Study Conditions
- Gas Condition
  - Oxygen vs Air
- Exercise Condition
  - Exercise vs Resting
- Immersion vs Air
- Dive

<u>499 Samples x 252 Variables</u>

Symptomatology (Total Sym)
- **Samples: 399**
- Class Balance: Sym/NoSym
  - 95/304 = **31.25%**



Immersion dives from the model basis, retaining dry conditions could confound results

Multiple Experimental Study Designs

Study Names

1. MULLET [2]
    - Study Focus: Hydration, Orthostatic tolerance, Exercise performance
        - MULLET X
        - MULLET
2. RIP [1]
    - Study Focus: Muscular strength and Exercise Performance
        - RIPX
        - RIP

Shared Similar Study Conditions

- Gas Condition
    - Oxygen vs Air
- Exercise Condition
    - Exercise vs Resting
- Dive
    - Single Dives and Dive Day (1-5)
- Pre vs Post Dive Physiology
- Immersion vs Air

General Study Variables

- Vascular/Endo
- Cardio
- Pulmonary
- Muscular
- Labs and Blood
- Demographics

[1] Myers, C.M, Kim Jeong-Su, Florian J.P. “Consecutive, Resting, Long-Duration Hyperoxic Exposures Alter Neuromuscular Responses During Maximal Strength Exercises in Trained Men.” Front. Physiology, Exercise Physiology Vol. 10 2019

[2] Considine EG, Florian JP, Klemp AO. “Endurance Exercise Performance Is Reduced after 6-h Dives at 1.35 ATA When Breathing 100% Oxygen Compared with Air.” Med Sci Sports Exercise. 2024 Feb 1;56(2):257-265.

Bottomline Upfront (BLUF)

- We merged numerous diving experimental data sets to test evaluate how multiple consecutive dives impose an accumulative physiology load impacting pulmonary symptomology (i.e., inflammation)

- Risk and threshold …

- Model B:
  - Pulmonary symptomatology appears to be driven by the balance between physiological reserve and accumulated inflammatory burden. Divers with greater baseline NO may possess greater airway resilience, allowing them to tolerate larger physiological perturbations following dive exposures. Conversely, symptom risk increases when residual inflammatory and physiological burden accumulates across consecutive diving days

Next Steps and Future Work

Modeling and Data Processing Perspective

- Leverage data fusion and uncertainty quantification to enhance predictive performance [1]
- Extend temporal coverage beyond pre- and post-dive points to capture dynamic physiological markers.

Outcomes Variable Perspective

- Expanding to additional performance limiting factors (e.g., exercise endurance) is a necessary and logical next step to ensure operational success

Utility Perspective

- Translate technologies from basic research to actionable TRLs to provide functional guidance to SOF operators

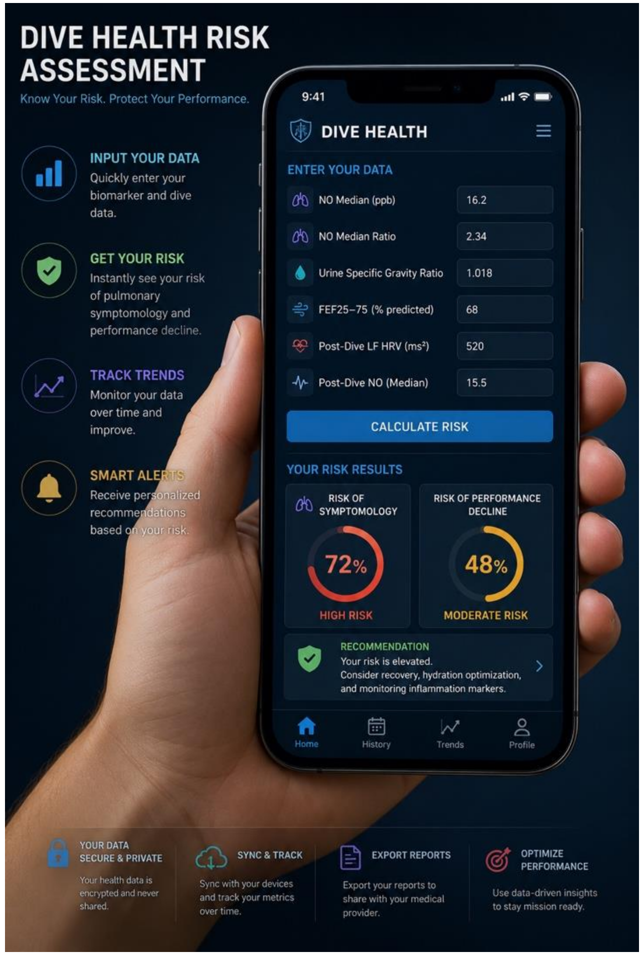


Photo was AI Generated using ChatGPT

[1] N. J. Napoli, C.L. Stephens, K.D. Kennedy, L.E. Barnes, E.J. Garcia, A.R. Harrivel, NAPS Fusion: A framework to overcome experimental data limitations to predict human performance and cognitive task outcomes, Information Fusion, 91, 2023, Pg 15-30