Penn Medicine

# Machine Learning Applications for Predicting ICU Stay Following Posterior Spinal Fusion in Adult Spinal Deformity

Mert Marcel Dagli, MD; Ali K Ozturk, MD; Jang W Yoon, MD; William C Welch, MD

June 27, 2024

# Disclosures

None

# Agenda

BACKGROUND

METHODS

RESULTS

CONCLUSIONS

# Background

Posterior spinal fusion surgery (PSF) for adult spinal deformity correction (ASD) demands intensive preoperative planning, including optimization predictions surrounding intensive care unit (ICU) utilization

Advanced machine learning (ML) algorithms have immense potential to support clinical decision-making and help with planning

Purpose: Development and Validation of predictive models using common machine learning approaches for ICU (length of stay) LOS following multi-level thoracolumbosacral PSF for ASD

# Methods

## PART 1

Guidelines: TRIPOD & STROBE

Data Source: retrospective review, institutional deformity database (DAC, charts)

Inclusion criteria: TLS PSF with 6+ vertebrae for ASD (n=646)

Data Cleaning and Handling of Missingness with Iterative Imputer with Random Forest and random states

Predictor Selection:

- Candidate predictors identified: literature, clinical importance, and analysis
- 5 predictors for ICU LOS ≤ 7d
- 10 predictors for ICU LOS ≤ 5d
- 13 predictors for ICU LOS ≥ 1d

# Methods

## PART 2

Data split: train-test-val 70-15-15

Models: Artificial Neural Network (ANN), ensemble boosted ANN (ebANN), extreme gradient boosting (XGBoost)

- Customization, grid search, regularization, fine-tuning, etc.

Integration ongoing: Support Vector Machine (SVM), Explainable Boosting Machine (EBM), Random Forest (RF), Logistic Regression (LR)

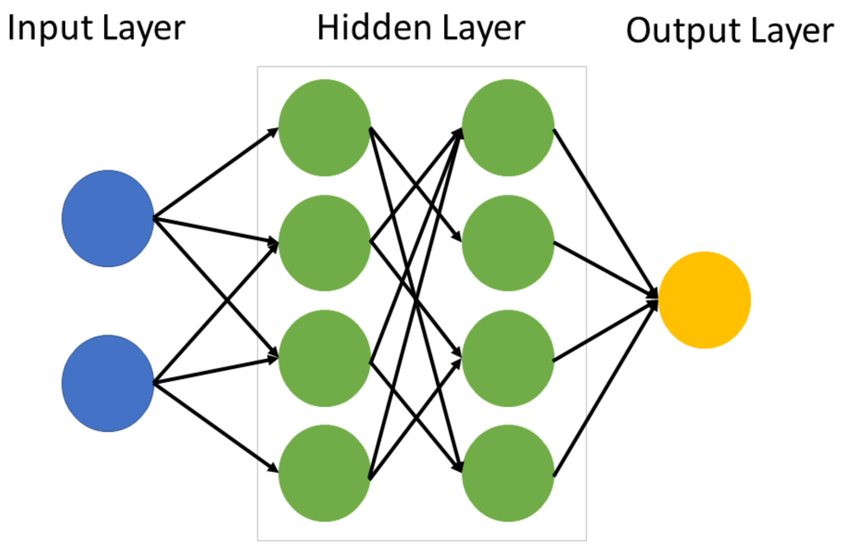
Performance metrices: accuracy, sensitivity, AUC-ROC (c-score), F1-score, precision (PPV) + Bootstrapping

# Methods

## PART 3 – MODEL ANN

ANN

- Package: PyTorch
- Standardization of features
- Batch size: 128
- Epochs: 200
- Patience: 20 (early stopping)
- Criterion: Binary Cross Entropy (BCE) loss
- Optimizer: AdamW (learning rate 0.01, weight decay 0.1)
- 4 layers: 2 layers input and output, 2 hidden layers, 12-128-64-1 with 20% and 10% dropout between layers 2-3 and 3-4



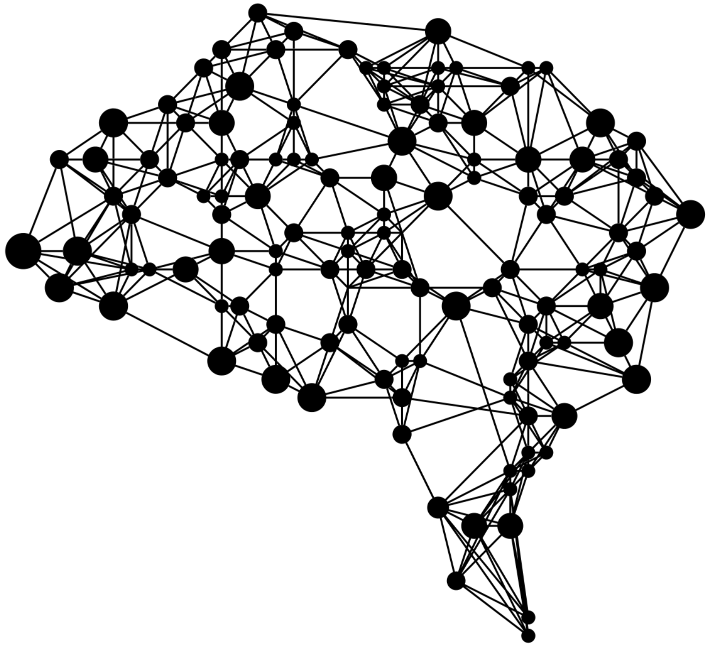

Artificial Neural Networks

# Methods

## PART 4 – MODEL EBANN

ebANN

- Package: PyTorch
- Ensemble ANNs with previous architecture
- Boosting rounds: 5 (iterative grid search range)

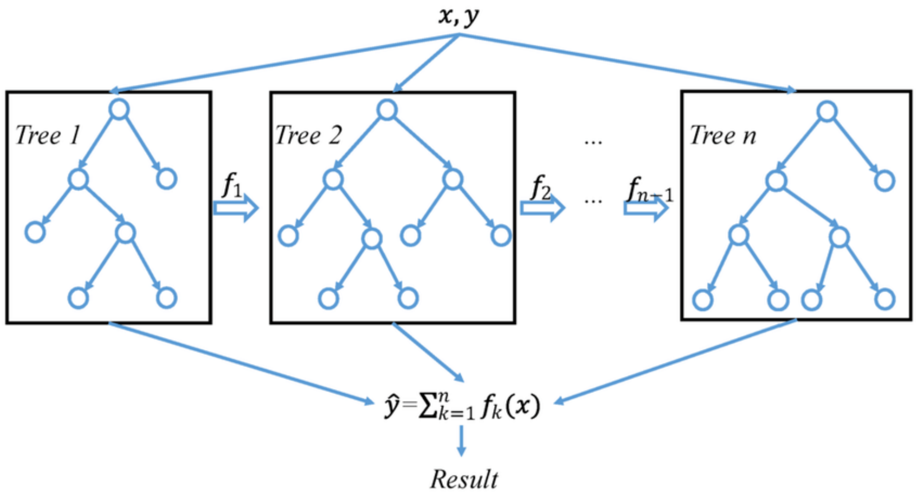


# Methods

## PART 4 – MODELS XGBOOST

XGBoost

- Max depth: 30
- Learning rate: 0.1
- Number of rounds: 1000
- Automatic early stopping (XGBoost)



# Methods

## STEP 5 – EVALUATION OF INTERNAL VALIDATION AND PERFORMANCE

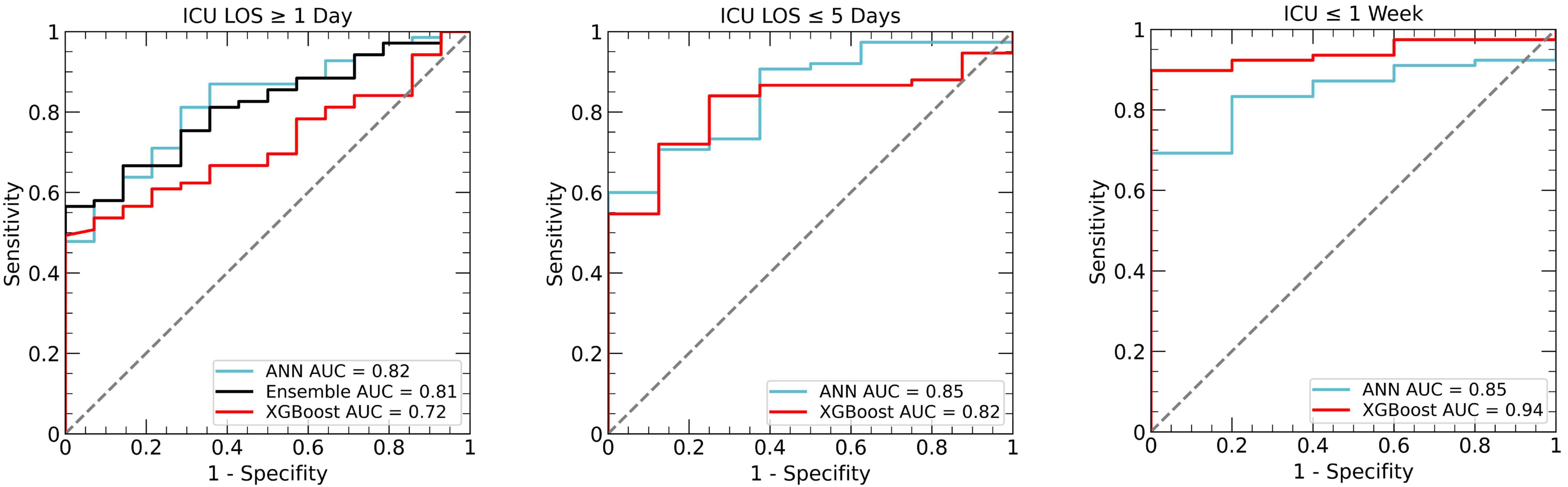
Evaluation of confusion matrix to evaluate predictions (TN, TP, FN, FP) is a form of internal validation.

Internal validation with testing set

Selection of standard performance metrices derived from testing set:

- Sensitivity: True Positive Rate
- Specificity: True Negative Rate
- Discrimination AUC-ROC: model's ability to discriminate between classes
- F1-Score: Harmonic mean between precision and sensitivity (minimizing FP, max. TP)
- Precision (PPV): ratio of TP predictions : all positive predictions

# Results


ICU LOS ≥ 1 Day
Sensitivity
1 - Specifity
ANN AUC = 0.82
Ensemble AUC = 0.81
XGBoost AUC = 0.72
ICU LOS ≤ 5 Days
ANN AUC = 0.85
XGBoost AUC = 0.82
ICU ≤ 1 Week
ANN AUC = 0.85
XGBoost AUC = 0.94

# Conclusions

A spectrum of explainable and opaque models to develop the best predictive model, while also including information on reasoning, is important and guides future research

Preliminary results of a small number of AI models for prediction of ICU LOS after multi-level TLS PSF for ASD was accurate and very promising

Integration of more AI models (explainable) and traditional LR is required to support our findings

Cut-offs of outcomes are being internally re-evaluated to maximize clinical importance (f.e. longer than 2 days)