Early Prognosis of Metabolic Dysfunction Associated Fatty Liver Disease using Deep Learning and Clinical Data Analysis

Jeannie She[1], Erica Song[2], Naiqi Zhang[3], Amy Tsurumi, PhD[4], Peter Szolovits, PhD[1]

[1]MIT Department of Electrical Engineering and Computer Science, [2]Harvard T.H. Chan School of Public Health, [3]MIT Sloan School of Management, [4]Massachusetts General Hospital

Early Prognosis of MAFLD Progression

Metabolic dysfunction associated fatty liver disease (MAFLD) is characterized by fat accumulation in the liver. Left undetected, it can advance to metabolic dysfunction-associated steatohepatitis, a more severe state that may result in cirrhosis or hepatocellular carcinoma.[1,4]

We train and evaluate deep learning models to predict a patient's prognosis in terms of **binary outcome, time-to-event, and survival modeling**. This extends prior work where regression models were trained to achieve AUROCs up to 0.80.[4]

Data Processing

Dataset: Demographics, diagnoses, labs, medications, and physical measurements from Mass General Brigham's Research Patient Data Repository as of January 2025.

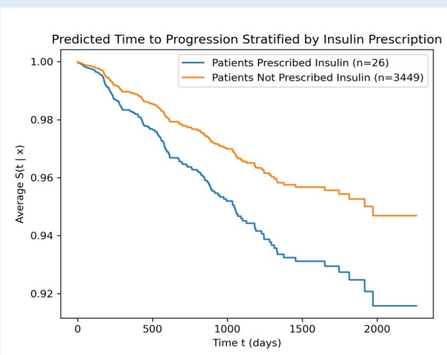
Cohort Selection (n = 11,582): Patients age 30 and above diagnosed with MASLD. Removed patients with worsened liver progression prior to MASLD diagnosis and alcohol- and drug-related disorders.

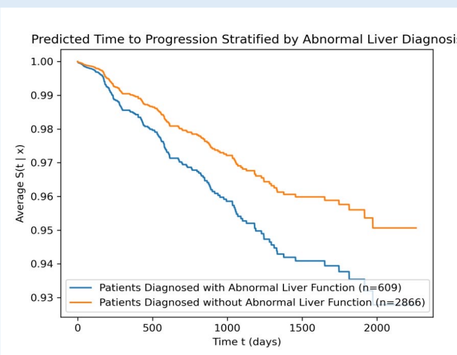
Feature Selection: Data within two years preceding first MASLD diagnosis. Feature extraction yielded 1,836 diagnoses, 89 lab values, 920 medications, and physical measurements such as BMI, with labs aggregated by mean value and BMI captured both as mean and most recent measurement before diagnosis.

Preprocessing: One-hot encoding of categorical variables and standardization of numerical variables.

Deep Learning vs. Baseline Model Results

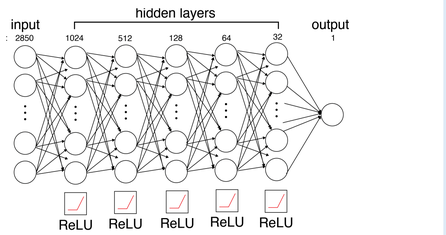
Deep Learning Models	Baseline Models
Binary Classification Neural Network Test AUROC = 0.516 Train AUROC = 1.00	**Logistic Regression** Test AUROC = 0.534 Train AUROC = 0.918
Time-to-Event Neural Network Test RMSE = 504 days Test MAE = 376 days	**Linear Regression** Test RMSE = 621 days Test MAE = 487 days
DeepSurv Model Test C-Index = 0.550 Test Integrated Brier Score = 0.0550	**Cox PH Model** Test C-Index = 0.594 Test Integrated Brier Score = 0.0496

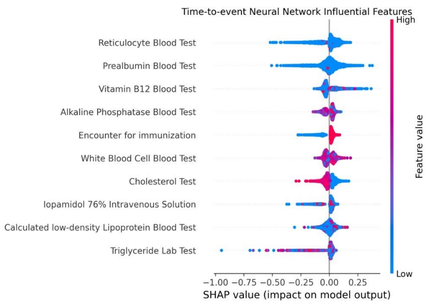




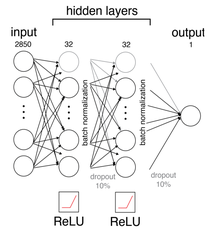
Deep Learning Architectures and SHAP Analysis

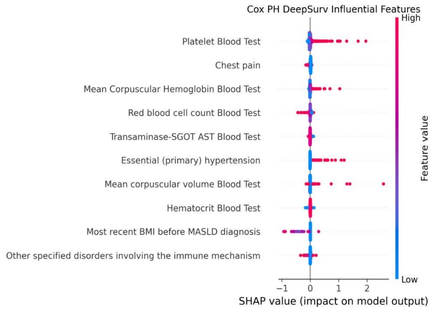
Time-to-Event Neural Network Architecture





DeepSurv Architecture





Cholesterol, LDL, and triglyceride blood tests, key lab tests for obesity patients, play a significant role in the time-to-event model output. AST blood test and last recorded BMI were key for survival modeling. Top influential features are consistent with known biomarkers for MAFLD.

Class Imbalance Strategies

SMOTE
Test AUROC = ~0.5
Synthetic samples of the minority class generated for training data by interpolating between existing samples and their nearest neighbors.

Downsampling
Average Test AUROC = ~0.5
Downsampled the majority class of patients who did not progress to match the number of patients who progressed. Repeated 10 times for robust training and averaged results.

Class Reweighting
Test AUROC = ~0.5
Weighted the loss function inversely proportional to class frequency with test AUROC=~0.5. Attempted 50:1 reweighting of minority class with test AUROC between 0.54 to 0.58.

Deep Learning Model Strengths and Limitations

Binary Classification Neural Network: Extreme overfitting.

Time-to-Event Neural Network: Strong performance relative to full study period.

DeepSurv Model: Good calibration and discrimination.

Limitations: No access to unstructured data, such as clinical notes. Severe class imbalance, such that SMOTE and reweighting offered limited improvement. No external data for validation. No access to specific date information.

Future Directions: Hyperparameter tuning and new model architectures. Leverage large language models (LLMs) for clinical note analysis. Investigate pediatric populations and broader outcomes.

[1] Metabolic Dysfunction-Associated Steatohepatitis: What It Is, Causes.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16:321–357, June 2002. arXiv:1106.1813 [cs].

[3] Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yu- val Kluger. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology, 18(1):24, February 2018.

[4] Jonathan Li. Predicting Progression of Metabolic Dysfunction-associated Steatotic Liver Dis- ease. Technical report, Massachusetts Institute of Technology, February 2024.

[5] Peng-Cheng Ma, Qi-Mei Li, Rui-Ning Li, Chang Hong, Hao Cui, Zi-Yong Zhang, Yan Li, Lu- Shan Xiao, Hong Zhu, Lin Zeng, Jun Xu, Wei-Nan Lai, and Li Liu. A high reticulocyte count is a risk factor for the onset of metabolic dysfunction-associated steatotic liver disease: Cross- sectional and prospective studies of data of 310,091 individuals from the UK Biobank. Frontiers in Pharmacology, 15, July 2024. Publisher: Frontiers.

[6] Leslie N. Smith. Cyclical Learning Rates for Training Neural Networks, April 2017. arXiv:1506.01186 [cs].