INST452
05/13/2025

HEPATITIS DETECTION

Samuel Muma, Chantel Orimoloye,

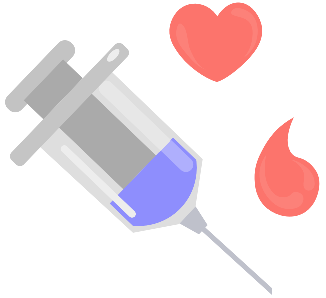
Dara Ojebouoh, Kyle Tabong

TABLE OF CONTENTS

TABLE OF CONTENTS

Can machine learning be leveraged to predict patient mortality due to hepatitis?

— Our Research Question



Background

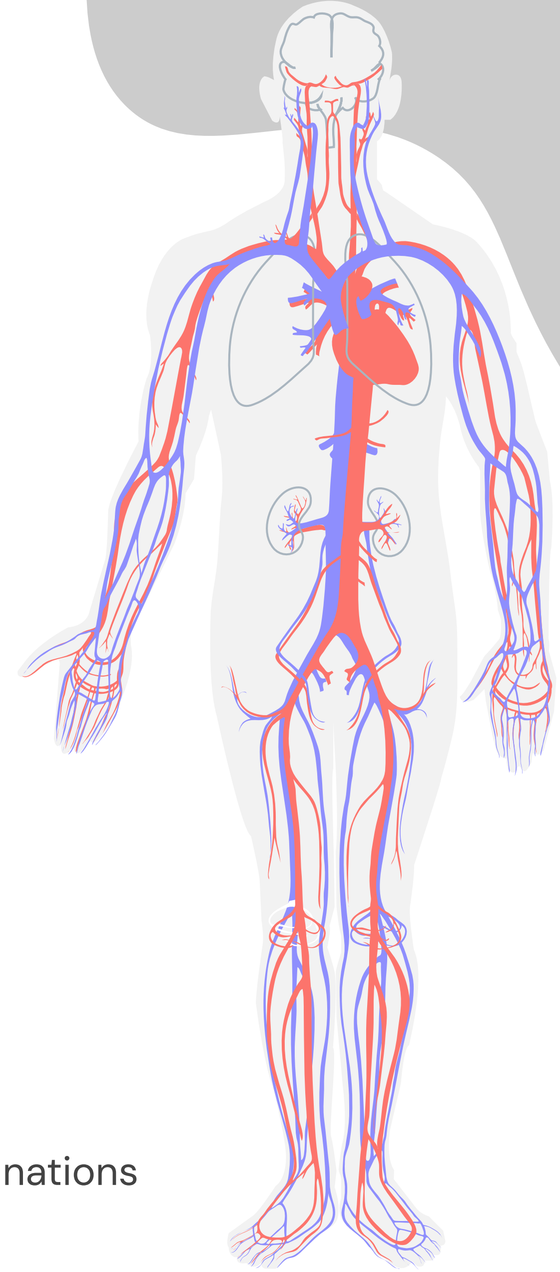
Hepatitis causes liver inflammation from viruses or other agents, leading to serious health risk if untreated.

Who benefits from this?

Clinicians: Helps identify high-risk patients quickly for timely interventions.

Researchers: Provides insights into hepatitis progress across demographics

Patients: Early diagnoses and personalized treatments



Potential Use Case:
Public Health organizations can create targeted vaccinations and preventative measures for high-risk groups.

DID YOU KNOW?

For children under 6

Less than 30% of young children show symptoms of Hepatitis A, meaning it often goes unnoticed in little ones.

For older children and adults

Over 70% of older children and adults will experience jaundice, a tell-tale sign of Hepatitis A.

Timeframe for symptoms

Hepatitis A symptoms can appear anywhere between 2 weeks to 6 months after exposure.

Very interesting facts!

Data Cleaning & Preprocessing

Missing Values

For binary features, the mode is imputed through a helper function, impute_mode.

Binary Features

Binary features are encoded to 0’s & 1’s for machine learning readability.

Continuous Features

Continuous features are imputed with either a mean or median through a helper function impute_with.

Target Feature

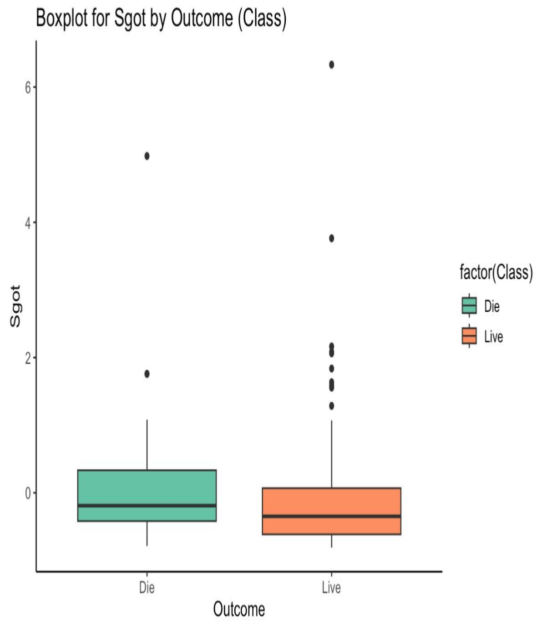
Our target feature was converted to a factor and labeled “Live/Die to assess outcomes..

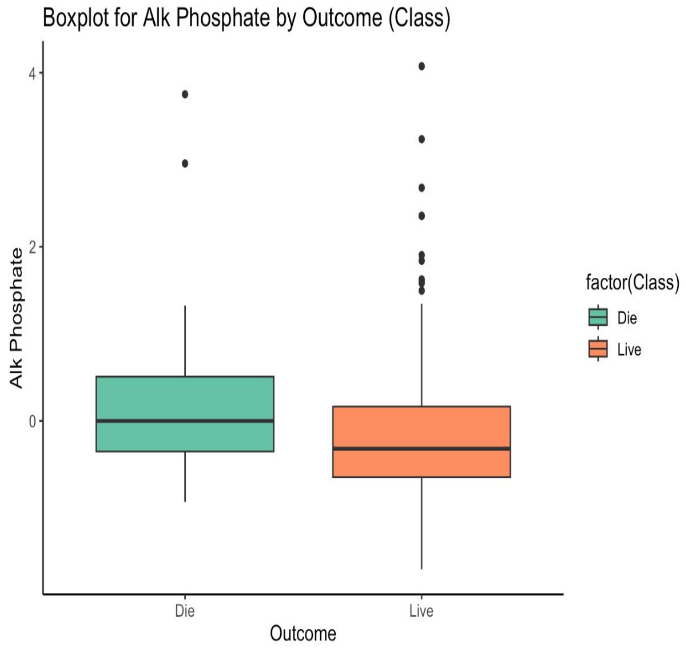
Scaling

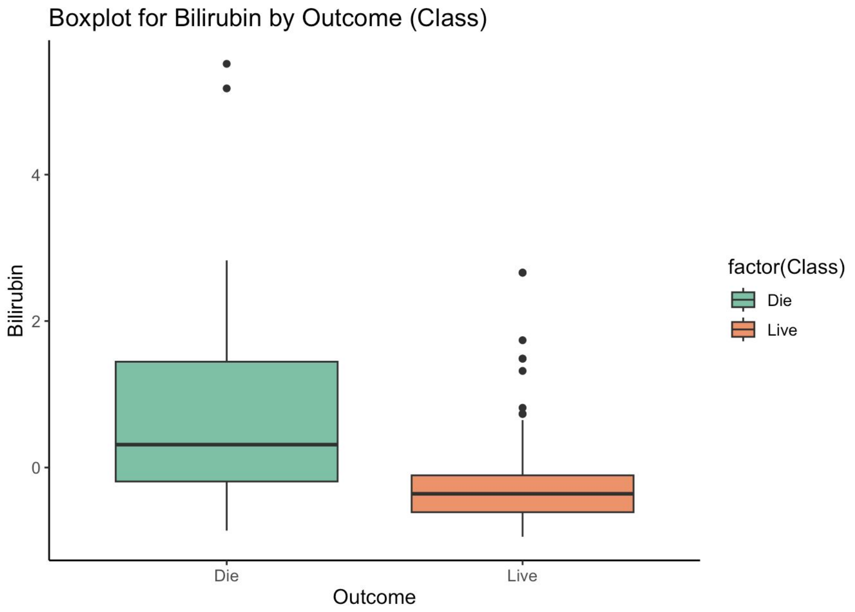
Continuous features were normalized through scaling for consistency and improved performance.

Dataset Source: UCI Repository. Content: 23 variables, 8257 observations

DATA VISUALIZATION







Insight:

Differences in distributions between "Live" and "Die" suggest these features help discriminate outcomes.

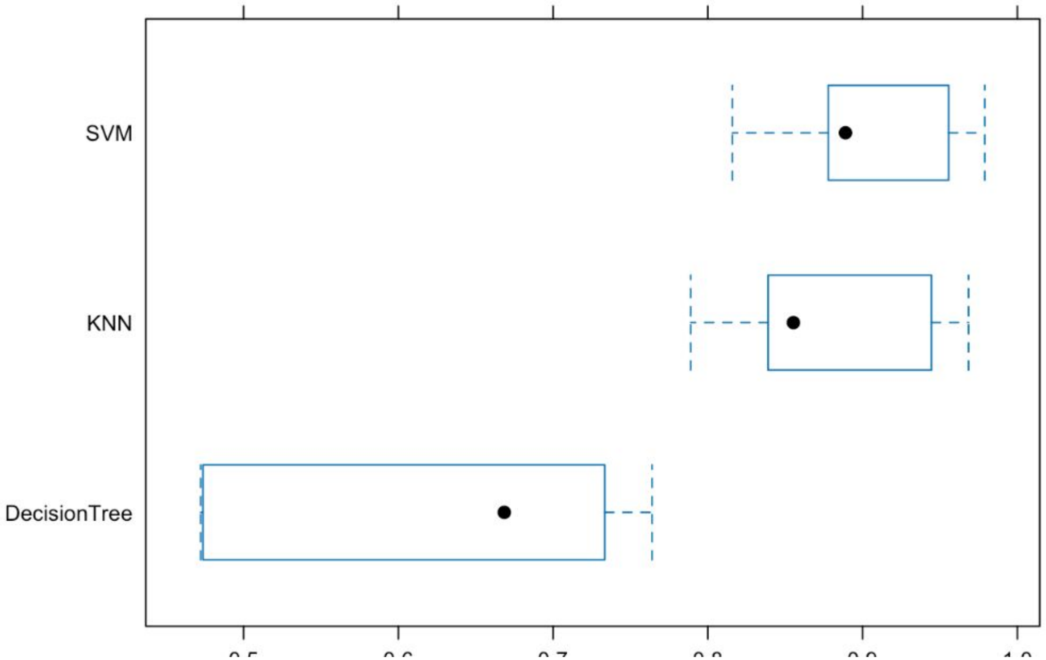
DATA VISUALIZATION

Bilirubin: Patients who died tend to have higher median Bilirubin levels, suggesting severe liver dysfunction.

Alk Phosphatase: Both groups have a similar spread, but the “Die” group shows slightly higher variability and higher median values, which may reflect liver stress.

Sgot: The “Die” group again shows higher median and wider spread, indicating worse liver enzyme levels.

CROSS MODEL AND PERFORMANCE



SVM shows the highest overall performance across all metrics, suggesting it may be the best candidate for predicting survival outcomes.

AUC values close to 1 show the models are effectively separating Live vs. Die cases.

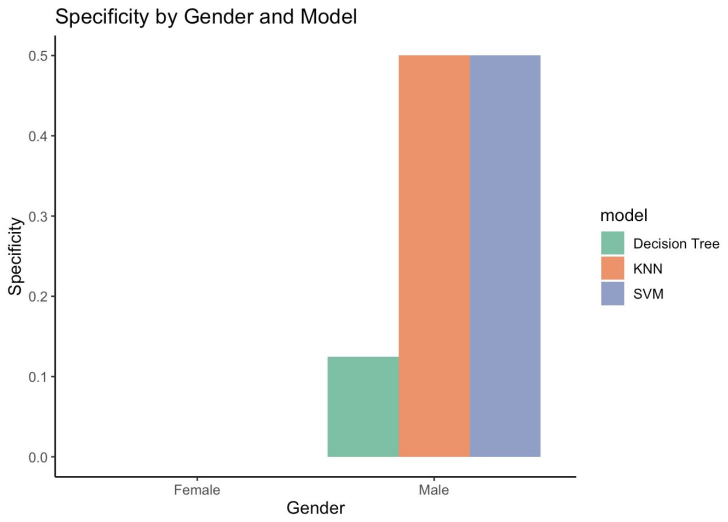
FAIRNESS & BIAS EVALUATION

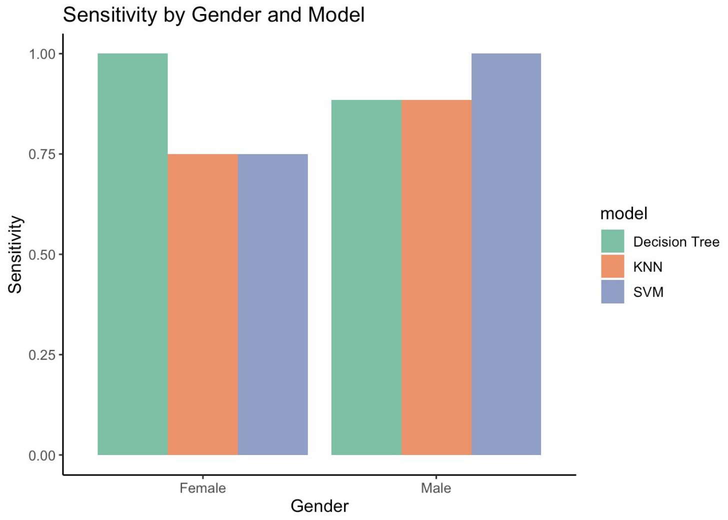
Models Compared: **Decision Tree, SVM, KNN (5-fold cross-validation)**

Key Metrics:

- **ROC/AUC:** Measures each model's ability to distinguish between classes.
- **Sensitivity:** True Positive (ability to predict Die)
- **Specificity:** True Negative (ability to predict Live)

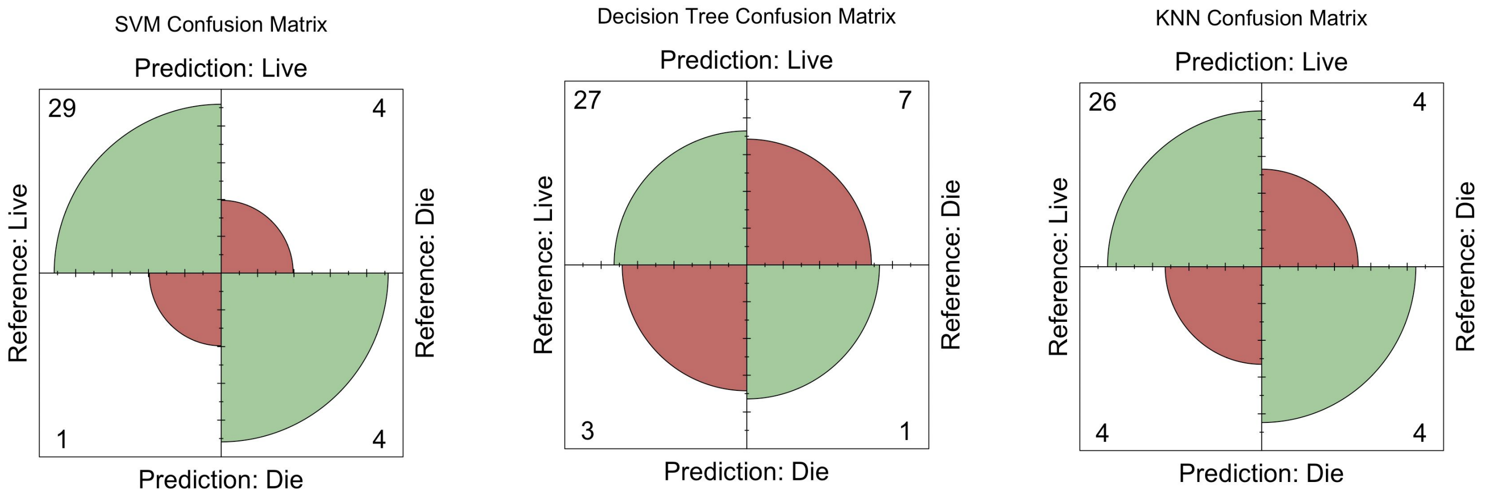
FAIRNESS & BIAS EVALUATION





We compared how well each model performed for male and female patients. The charts show how often the models correctly predicted survival (sensitivity) and avoided false predictions of death (specificity). This helps us see if the models work equally well for both groups or if one group is treated less fairly. Ensuring fairness is important for making reliable clinical decisions.

TEST SET EVALUATION



Each chart shows how well the models predicted patient outcomes (“Live” or “Die”). The top left and bottom right quadrants represent correct predictions. The SVM model had the best balance—accurately identifying both survivors and those who died. The Decision Tree misclassified more patients who died, while KNN showed more false positives. These matrices help us evaluate accuracy and spot potential clinical risks in each model’s predictions.

METRIC SELECTION

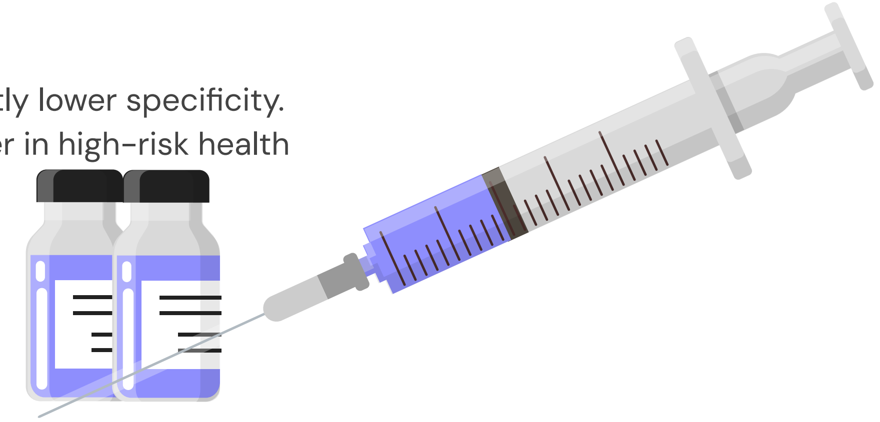
Our Selected Metric: **Sensitivity (Recall)**

Why Prioritize Sensitivity?

- In clinical decision-making, missing a patient who might survive is far more harmful than a false positive.
- Sensitivity focuses on correctly identifying actual survivors (those who could benefit from treatment or intervention).
- High sensitivity reduces the risk of undertreating or overlooking patients who still have a chance of recovery.
- False negatives (misclassifying a patient as "Die" when they could "Live") could deny life-saving treatments, which is a serious ethical and medical risk.

Real-World Implication:

- Prioritizing sensitivity helps save lives, even if it means accepting a slightly lower specificity.
- It ensures more patients are flagged for care or monitoring, which is safer in high-risk health conditions like hepatitis.



RECOMMENDATIONS TO CLINICIANS

Based on the findings the SVM will be recommended for clinical use

- Best Sensitivity (85%) among all models, meaning it is the least likely to miss survivors.
- Strongest overall Accuracy (82%) and AUC (0.84), showing it performs well across multiple metrics.
- Balanced Specificity (78%), meaning it still limits unnecessary treatment for patients who are unlikely to survive.

From a clinical Perspective

- SVM provides the safest balance between identifying those who can be saved and minimizing false alarms.
- Helps clinicians triage patients more effectively by focusing on those with the highest survival likelihood.

Improvements that can be made

- The dataset could be more diverse to improve generalizability, which is more consistent.
- Include more clinical features (e.g., lifestyle, medication history) to improve prediction.

CITATIONS.

- Presentation Template: SlidesMania
- Images: Unsplash
- Fonts used in this presentation: DM Sans and Unica One
- WHO. (n.d.-b). *Hepatitis*. World Health Organization. https://www.who.int/health-topics/hepatitis#tab=tab_1
- Centers for Disease Control and Prevention. (n.d.). *Clinical overview of viral hepatitis*. Centers for Disease Control and Prevention. https://www.cdc.gov/hepatitis/hcp/clinical-overview/index.html

Thank you!

Do you have any questions?