Team 4
DSO 568
Maureen Dzifa Awumee Quist, Chelsea Baker, Karen Botros, Elmer Camargo, Uchechukwu Ezealigo, Sam Sithimolada

Sepsis Survival Prediction Model

Executive Summary:

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection commonly associated with bacterial infections, but can also result from viral, fungal, or parasitic infections. It represents a major global health burden, contributing to high mortality and morbidity rates in both high and low-resource settings. Approximately 1.7 million cases of sepsis occur annually in the U.S., with mortality rates ranging from 15-30%. The prognosis and survival of this condition depend on timely recognition and intervention.

Preventing the contraction of sepsis is extremely important. However, if this is unlikely, predicting the survival of patients as quickly as possible is equally as important. Sepsis spreads quickly throughout the body, causing severe organ damage. Although lab tests are important for diagnosing and understanding the causes of sepsis, they tend to take a long time to process and give results. This delay can make the difference between survival and death, as every minute counts when treating sepsis. Prompt and accurate predictions could significantly improve the outcome of the patient by enabling faster decision-making and treatment.

Topic and Problem Definition:

Sepsis is a disease that occurs when an infection in the body reaches the blood and the immune system has an extreme reaction. According to the CDC, sepsis is the leading cause of death in hospitalized patients in the United States. In Norway, where the dataset comes from, it is the leading cause of death from infection. This reaction can be life-threatening and impacts the entire body as it spreads. It is important to catch the symptoms as early as possible and treat them immediately to increase the likelihood of saving the patient's life. This is especially true considering how sepsis can kill patients within as little as twelve hours. Despite advances in medical technology, sepsis remains a significant public health challenge due to how rapidly it spreads and the complexity of early detection. This project looks at sepsis cases in Norway and how sepsis-related hospitalizations contribute to overall hospital mortality.

Purpose:

When diagnosing and treating a sepsis case urgency is of the utmost importance, as even a delay of an hour could make the difference between a patient living and dying. Because of this, identifying risk factors and the outcomes associated with sepsis-related hospitalizations is important for improving patient care, outcomes, and the overall mortality rate of hospitals. By analyzing this data, we may find patterns and predictors of sepsis outcomes and identify the next steps for healthcare providers to ensure that they provide the best care possible. We can also help shape healthcare policies and protocols to aid sepsis prevention, early recognition, and treatment strategies, which can improve survival rates for hospitalized patients in Norway and other countries. By reducing sepsis-related admittances and mortality rates, we can also lower healthcare costs associated with prolonged hospital stays, readmissions, and intensive care unit treatments. Early identification and management of sepsis can improve hospital efficiency, optimize resource allocation, and reduce the financial strain on healthcare systems.

Data Overview:

The data comes from hospitals in Norway between 2011 and 2012, from patients who have sepsis-related infections. It has 110,204 data points. The data set has five main variables:

- **Age (years)**: Age of the patient
    - Older patients may have weaker immune systems, making them more susceptible to complications and higher mortality rates from sepsis.
- **Sex (0 male, 1 female):** Sex of the patient
    - Biological differences in immune response between males and females may influence sepsis outcomes.
- **Length of Stay (days)**: How long they have been in the hospital for
    - The longer the stay at the hospital, the more severe the sepsis or complications are, increasing the risk of death.
- **Hospital Outcome (0 alive, 1 dead):** The outcome of the patient's stay after about nine days of their medical record being collected
- **Episode Number**: The number of times they have been admitted for sepsis-related condition
    - If a patient has been hospitalized more than once for sepsis-related conditions, they may be more vulnerable, impacting their risk of dying.

There are nine ICD (International Classification of Disease) codes, which code diseases and conditions. There are some missing values for the ICD codes, so we dropped them.

Data Preparation and Cleaning:

To preprocess and prepare the data we checked for missing values. We found 18 missing values, all of which were missing ICD-10 codes to identify which diseases the patient had. Since our data set has 110,000 total entries, we decided to drop these values because they are negligible compared to the rest of the complete data. We added up the number of conditions for each patient and then dropped the ICD-10 codes to simplify analyzing the data.

Introductory Analysis:

We did some introductory graphs and analysis to understand the data before we built the models. Some summary statistics we found are:

- There are 57,973 male patients and 52,231 female patients
- The average patient age is 63 years old
- The average length of stay is 9.35 days
- The average episode number of a patient is 1.35
- Out of the entire data set, 102,085 patients survived, and 8,103 died

Model Building:

The models we have chosen to build include logistic regression, random forests, decision trees, bagging classifiers, and neural networks. One significant limitation of our dataset is the class imbalance, where only 7.35% of the data corresponds to "dead" cases, compared to 92.65% "alive" cases. This imbalance poses a challenge to our model's performance, particularly in recall, which is a critical metric in this scenario. A low recall is concerning because it indicates that the model may fail to identify high-risk patients, leading to missed opportunities for timely, life-saving interventions. Given the high stakes in predicting patient outcomes, prioritizing recall is essential to ensure that as many at-risk patients as possible are accurately identified for appropriate treatment. To address this limitation, we applied balanced class weights, which, while improving recall, reduced accuracy.

Model Evaluation:

The tables below showcase the performance of our models, comparing results both with and without the application of balanced class weights. To ensure an equitable evaluation of metrics such as accuracy, precision, recall, and F1-score, a weighted average was employed during calculations. This approach ensures that the overall performance score reflects the relative contribution of each class, accounting for the inherent class imbalance in the dataset.

Balanced class weights, when applied, shift focus to underrepresented classes, improving the ability to detect minority (dead) cases without overly compromising performance for the majority class. These adjusted scores are critical for addressing the imbalance and ensuring fair model evaluation in high-stakes scenarios.

Training performance comparison:

	Logistic Regression	Logistic Regression - weighted class	Decision Tree	Decision Tree - weighted class	Decision Tree Tuned	Random Forest	Random Forest - weighted class	Bagging Classifier	Bagging Estimator Tuned	Neural Networks
Accuracy	0.925899	0.658371	0.927328	0.603530	0.927328	0.949582	0.870558	0.946213	0.937638	0.645826
Recall	0.925899	0.733210	0.927328	0.792052	0.927328	0.949582	0.870558	0.946213	0.937638	0.645826
Precision	0.889571	0.143603	0.903569	0.132710	0.903569	0.946525	0.936752	0.940762	0.938933	0.912224
F1	0.892866	0.240168	0.897971	0.227331	0.897971	0.939706	0.893398	0.935761	0.916380	0.730092

Testing performance comparison:

	Logistic Regression	Logistic Regression - weighted class	Decision Tree	Decision Tree - weighted class	Decision Tree Tuned	Random Forest	Random Forest - weighted class	Bagging Classifier	Bagging Estimator Tuned	Neural Networks
Accuracy	0.926228	0.657502	0.928542	0.601289	0.928542	0.918243	0.817295	0.915839	0.927181	0.642122
Recall	0.926228	0.729076	0.928542	0.792932	0.928542	0.918243	0.817295	0.915839	0.927181	0.642122
Precision	0.885548	0.141892	0.908368	0.131409	0.908368	0.887639	0.881153	0.885871	0.900618	0.911810
F1	0.892708	0.237552	0.900128	0.225454	0.900128	0.898277	0.845495	0.897109	0.899244	0.727458

Model Selection:

The Bagging Estimator Tuned model is the best overall, as it achieves the highest metrics on both training and testing datasets while demonstrating minimal overfitting (its training and testing metrics are closely aligned). Since recall is a critical metric for this prediction model, the Bagging Estimator Tuned model's superior recall performance further solidifies its suitability.

Insights:

As noted, our most performant model was a bagging classifier using tuned *sklearn* settings. An issue we came across with this model is that the tuning possibly included randomization on what variables assigned for our algorithm making it difficult to single out importance scores. In lieu of explanation on the most performant model we offer insight into the second most performant model that performed very similarly - default bagging classifier. The way the model works is that a base estimator of 10 Decision Trees is instantiated with the default setting of using gini impurity as the criterion for splits. During the training process, a feature importance score is derived for each of our 5 features - Age, Sex, Length of Stay, Episode Number and Number of Medical Conditions; the process is visualized in Fig A in the appendix. The average of all importance scores is calculated within our bagging classifier and visualized in Fig B. **The features playing the largest roles across all models are shown to be Age and Length of Stay**. These two variables are major determinants of healthcare outcomes for patients experiencing sepsis. It is important to note that determining how exactly the variables play into predicting the outcome cannot be easily interpreted using bagging and tree methods. A note on potential ways to evaluate interpretation is offered in the future work section, however, the two variables typically play a large role in survival for inpatient acute medical conditions.

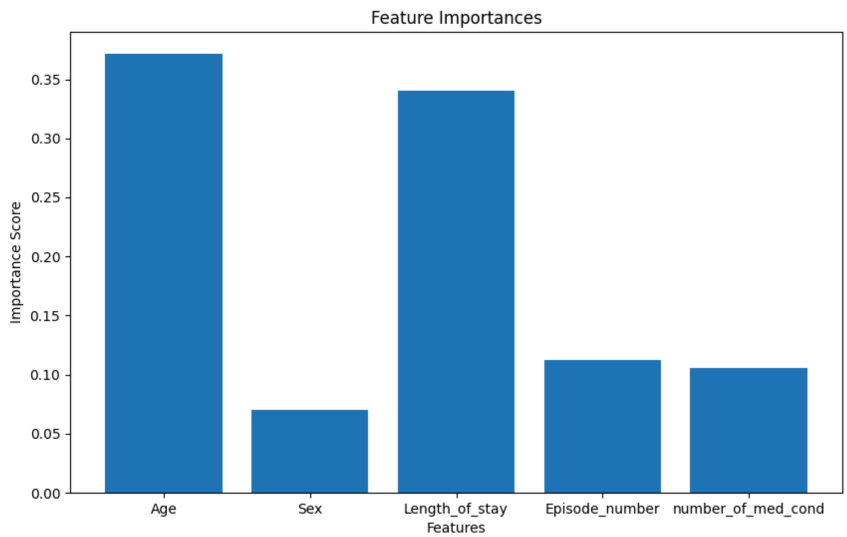

Figure B - Feature importance chart for proposed bagging classifier

Limitations:

A limitation of our data is that it was conducted in Norway, and their healthcare system differs from that of the United States and other countries. Norway has universal healthcare, allowing everyone to access good quality healthcare services without going through financial hardship. While the insights found are helpful, the actions that can be taken to help improve patient outcomes may not be possible to implement in the U.S. due to healthcare policies and government regulations.

Another limitation is the lack of diversity in the data. Since the data comes from Norway, the data reflects and has a bias toward Norwegian patients and their demographics. This bias does not represent a more diverse population, particularly in the U.S., which could limit how generalizable the insights we found are to other countries. Other populations may have more prevalent risk factors and health issues, as well as differing access to healthcare, which is not accounted for in this data.

Business Implications For Healthcare Providers

From the tuning bagging model, there are pros and cons.

Success

1. The prediction of survival is high. About 92.30% (20,343 persons) of the sample. There is an efficient use of resources. By accurately predicting survival, hospitals can allocate critical resources such as advanced treatment, ICU beds to patients who are more likely to benefit, thereby enhancing operational efficiency and patient satisfaction.
2. Prediction of non-survival revealed that about 93 affected persons will be casualties with as low as 0.42%. This minimizes aggressive care as the patient may die regardless of the resources invested. However, this could raise ethical dilemmas and financial liabilities related to denying treatment to patients.

Failure

1. Predicted will be alive but dead 6.9% - 1,520 persons. Patients initially predicted to survive but died. This case represents wasted resources or investments in care that did not yield successful outcomes, which can drive up cost per patient and reduce overall profitability.
2. Predicted will not survive but lived 0.39% - 85 persons. This case results in the loss of revenue for the hospital or pay for a value based fee for that patient. Patients initially predicted not to survive (will die) but did survive may not receive aggressive care. Patient's survival without adequate intervention raises concerns about negligence or suboptimal care delivery and financial penalties under the value-base model.

Clinical Recommendations

Healthcare Providers:

1. Create awareness through public education and strengthen early detection and monitoring protocols (such as assay tool kits) especially for those with the top comorbidities for sepsis.
2. Enhance resource allocation. Establish "safety net protocols" to prevent neglect of patients predicted not to survive but who may also benefit from intensive care.
3. Incorporate this data analysis findings into providing improved quality initiatives by training staff. Also, patient-centered medical homes should be supported to handle special cases, particularly the male cases due to the low survival rate.

Academic Research, Biotechnology and Pharmaceuticals

1. Optimize clinical trials with focus on false negative groups (predicrted to survive but did not) to develop drugs addressing the unmet clinical needs.
2. Invest in early-stage intervention drugs for patients with hig-risk models.
3. Develop value-based pricing models for therapies, ensure pricing aligns with real-world patient outcome.

Finally, the government and orgnaizations should contribute to the funding of advance research on sepsis.

Future Work and Ethical Considerations:

There are a number of considerations regarding how to employ our suggested model in hospital contexts. For purposes of our research, we suggest adopting the privacy and ethics framework developed by the Department of Health and Human Services (DHH) for usage in the US. The privacy and security framework covers important characteristics that are generally considered best practices in healthcare; though every country may have different laws and regulations and research into specific country guidelines is important.

Assuming that our model is employed at a hospital in the US, the processes that involve this sepsis dataset or a hospital's own local dataset, e.g. our models or future models, should adhere to the framework components. Three components, or principles, of note that directly impact the operationalization of our model are noted below in Figure C. Other principles still apply, however, may fall more into the responsibility of cybersecurity, cloud operations, customer service, and various other departments. The headliner statements for these principles are vital to any venture within a healthcare space and are the following.

Privacy and Ethics Component	**DHH Statement**
Openness and Transparency	There should be openness and transparency about policies, procedures, and technologies that directly affect individuals and/or their individually identifiable health information.
Individual Choice Principle	Individuals should be provided a reasonable opportunity and capability to make informed decisions about the collection, use, and disclosure of their individually identifiable health information.
Collection, Use, and Limited Disclosure	Individually identifiable health information should be collected, used, and/or disclosed

	only to the extent necessary to accomplish a specified purpose(s) and never to discriminate inappropriately.

Figure C - Relevant DHH Privacy and Ethics Principles for Data Management and Operations

An example of how a hospital can employ the three noted principles is by considering and investigating the three following points alongside a trial period. (1) provide opt-(in/out) choices to patients, (2) limit/expand modeling features and training records, explore the explainable artificial intelligence (XAI) field and related tools. Notifying patients at inpatient check-in when being hospitalized, or ideally before, is an opportune time to make them aware of how their information is used and any processes related to AI-assisted diagnosis and prediction. Provided that a hospital adopts this model, they should take necessary steps to adjust their tech infrastructure to accommodate retraining procedures and supplement the provided research data with their own patient datasets and modeling methodologies. In other words, hospitals may need to work with their electronic medical records (EMR) and data systems to capture necessary information for our baseline model. Hospitals should then also be able to use their own data with possibly more patient details (or less) for future models at the discretion of data teams and privacy and ethics principles keeping the limited disclosure component in mind.

Lastly, an additional improvement that a hospital can make in implementing this model is to consider including a transparency layer in the modeling process. One tool gaining more traction in the healthcare and AI space is Shapely Additive Explanations (SHAP) which does as mentioned – act as a transparency layer that is agnostic to modeling algorithms. Cited from the authors – SHAP applies a game theoretic approach to explain the output of any machine learning model and allows for quantified contribution of each feature to individual predictions. It is in the best interest of any healthcare professional entity to consider the adoption of this technology, or similar technology, into their model development practice. Our recommendation, in addition to integrating the proposed model into a hospital healthcare setting, is to also adopt the mentioned DHH privacy framework and use our provided example as a use case to consider.

Conclusion:

Sepsis-survival prediction modeling is extremely important. Not only can we save lives, we can also lower associated healthcare costs. By analyzing the data, we found that age and length of stay are the most significant predictors of survival. Out of all the models, the tuned Bagging Estimator performed the best when predicting the outcomes of patients. It had the highest recall, which is crucial to correctly predicting patient survival. Although we were limited to a dataset from Norway, our insights can be applied to the U.S. healthcare system because sepsis is a prevalent issue. Future work should focus on adapting these models to other healthcare contexts, ensuring diversity in training data, and incorporating transparency frameworks like SHAP for better explainability. By incorporating our recommendations, healthcare systems can improve the quality of care they provide.

Appendix:

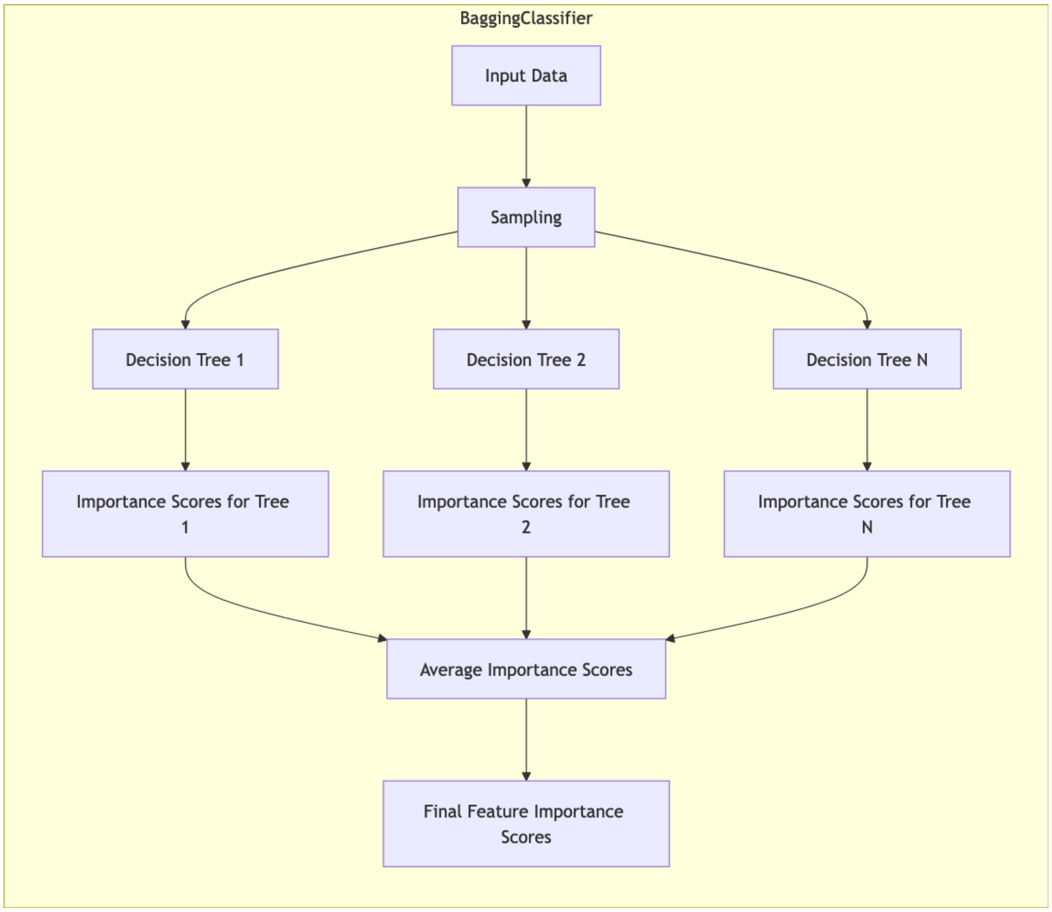

Figure A - Feature importance calculation for bagging classifier