

**Deep Learning**

**Predict 90-days mortality for MIMIC-III patients with Septicemia**

***Shivadath Gannavarapu***

**School of Graduate Professional Studies**

Data Analytics

DAAN 570 – Deep Learning

*Spring Semester, 2023*

**Document Control**

**Work carried out by:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Email Address** | **Exhaustive list of Tasks** |
| Shivadath Gannavarapu | [Szg6021@psu.edu](mailto:Szg6021@psu.edu) |  |
|  |  |  |
|  |  |  |

**Revision Sheet**

|  |  |
| --- | --- |
| **Date** | **Revision Description** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Contents

[1 Introduction 3](#_Toc128268580)

[2 Problem Statement 5](#_Toc128268581)

[3 Challenges 5](#_Toc128268582)

[3.1 RELATED WORKS 7](#_Toc128268583)

[3.2 IMPORTANCE AND IMPACTS 7](#_Toc128268584)

[4 Data Collection 8](#_Toc128268585)

[5 Data Preprocessing 11](#_Toc128268586)

[6 Methodology 22](#_Toc128268587)

[7 Results and INterpretation 25](#_Toc128268588)

[8 Discussion of Results 31](#_Toc128268589)

[9 Your Feedback 32](#_Toc128268590)

[10 References 32](#_Toc128268591)

# Introduction

Septicemia, also known as sepsis, is a life-threatening medical condition that occurs when an infection spreads throughout the body and triggers a systemic inflammatory response. This severe condition can lead to multiple organ failures and is one of the leading causes of death among hospital patients. Early detection and prompt treatment of septicemia are critical for improving patient outcomes and reducing mortality rates. In this project, we aim to develop a deep learning model to predict 90-day mortality for patients with septicemia using the MIMIC-III dataset.

The body's immune system is a highly developed response to infections that can be caused by bacteria, viruses, or fungi. However, when the immune system is unable to mount a tailored defense against infection, it releases an avalanche of inflammatory chemicals in order to create a mass effect, which leads to a state of sepsis within the body [1].

Sepsis is considered to be a severe and challenging condition that is characterized by a range of symptoms and causes widespread organ failure. It is often difficult to diagnose and treat, making it a lethal condition. Currently, it is defined as a condition that results from a disordered host response to an infection, leading to lethal organ dysfunction [2]. The primary cause of organ damage in sepsis is the discrepancy between the tissue's metabolic requirements and the decreased blood flow that results from the body's overactive immune response. This problem is compounded by the fact that inflammation causes cardiac problems and redistributes blood flow, while the damaged tissue also requires more oxygen, leading to a vicious cycle of organ failure [3].

Angiopoietins are a subset of a family of vascular growth factors. The imbalance of angiopoietin-tyrosine kinase alongside immunoglobulin-like ligand-receptor system (Ang-tie), which is responsible for cardiovascular and lymphatic development, is of particular interest in sepsis research [2]. The improved expression of Ang-2 and the impediment of Ang-1 obstructs the Tie-2 receptor and proliferates vascular permeability, causing tissue edema [4]. A high serum Ang-2/Ang-1 ratio in turn results in heightened severity in organ malfunction and increased mortality, even in early sepsis [5]. The organs often damaged due to sepsis include the kidneys, lungs, liver, heart, central nervous hematologic systems [2].

The infections that contribute to sepsis are resistant to antibiotics, leading to quick deterioration of health conditions [6]. The symptoms of sepsis include fever, irregular heart rate, blood vessel leaks, inflammation and clotting difficulties [7]. Sepsis can be classified as sepsis 1, 2 or 3. Sepsis 1 is known to occur as a consequence of the systemic inflammatory response to known or suspected infections. Sepsis 2 is diagnosed based on inflammatory, hemodynamic, organ maladies and tissue perfusion parameters, while sepsis 3 is diagnosed based on life-threatening malady of the organ due to the imbalanced host response to infection [8]. Figure 1 encapsulates the causes and consequences of sepsis. Sepsis afflicts more than 30 million people globally, causing about 6 million deaths yearly [9].

Early detection and timely management of sepsis are crucial to lower the mortality and morbidity rates. Presently, blood cultures are examined and biomarkers such as procalcitonin(PCT), C-reactive protein(CRP), cell-free DNA(cfDNA) are used as the gold standard for early sepsis diagnosis [10,11]. Despite the widespread use of blood cultures, this diagnostic approach has limitations, such as invasiveness. Additionally, the current biomarkers used for sepsis diagnosis, such as procalcitonin, C-reactive protein, and cell-free DNA, have been shown to be less effective in detecting sepsis due to their overlapping nature with other infectious and inflammatory conditions. [12]. The need for an affordable, automated diagnostic tool to detect sepsis early on is crucial because it can reduce the time for advanced diagnostics and allow for prompt treatment. However, existing Multimarker systems that aim to overcome the shortcomings of traditional diagnostic tools are costly and difficult to incorporate into clinical practices [13].

Diagram

Description automatically generated

# Problem Statement

Early detection of septicemia is critical for effective treatment and reducing mortality rates. This serious condition can cause multiple organ failures and is a major health concern. Despite the development of traditional statistical models for predicting septicemia, deep learning models have the ability to offer a more comprehensive and accurate analysis of patient data.

Till this day, Sepsis exhibits a high mortality rate. The multifactorial characteristic of this disease makes early diagnosis a challenging task to doctors and physicians. Additionally, the definition of sepsis exhibits a low specificity resulting in many patients that are wrongly identified as manifesting sepsis. In 1991, the first definition of sepsis and its different severity levels - severe sepsis and septic shock - was developed [16]. This definition was extended in 2001 to facilitate the bedside diagnosis of sepsis. Finally, in 2016 the whole definition was renewed in order to clarify the state of sepsis and therefore to facilitate earlier recognition of sepsis [17]. Nevertheless, the definition of 2016 is criticized for its potential of leading to higher mortality due to the downgrading of the sepsis definition to infection and severe sepsis to sepsis. As per the medical reports it says that even after discharging from the hospital, it is mainly observed that many people are dying within 90 days from the day of discharge.

In this project, we aim to use the MIMIC-III dataset to build a deep learning model that can predict 90-day mortality for patients with septicemia. The database we use for our retrospective analysis, the Medical Information Mart for Intensive Care (MIMIC) III database [18], was recorded between 2001 and 2012.

# Challenges

Forecasting sepsis is a complicated process that demands consideration of multiple variables, such as demographic information, medical history, and laboratory results. This can be difficult to handle due to a vast number of features and the inconsistent nature of the data. Furthermore, the MIMIC-III database is extensive and intricate, making it hard to process and prepare the data.

**Data Privacy:** The MIMIC-III dataset contains sensitive patient data, which means that it is important to protect the privacy of the patients.

**Data Quality**: The quality of the data in the MIMIC-III dataset is not always consistent, and it can be difficult to determine which data is reliable. This can result in biased predictions and decreased accuracy.

**Data Cleaning**: The MIMIC-III dataset is large and complex, which means that it requires extensive cleaning and preprocessing. This includes removing missing data, dealing with outliers, and transforming variables and merging the data frames to get a final data frame for analysis.

**Data Imbalance:** The number of patients in the MIMIC-III dataset with septicemia is much smaller than the number of patients without septicemia. This can result in a biased prediction model that only predicts mortality for a small number of patients.

**High-dimensional Data:** The MIMIC-III dataset contains a large number of features, which can make it difficult for deep learning models to extract meaningful patterns in the data.

**Limited Lab Data:** The MIMIC-III dataset only contains laboratory data for a small subset of patients. This can result in missing information that is important for making accurate predictions.

**Overfitting:** Overfitting can occur when a deep learning model is trained on too much data, resulting in a model that is too complex for the problem at hand.

**Time-series Data**: The MIMIC-III dataset contains time-series data, which means that the data for each patient changes over time. This can make it difficult for deep learning models to extract meaningful patterns in the data.

**Model Selection:** Selecting the appropriate deep learning model is a challenging task. There are many different models to choose from, each with its own strengths and weaknesses.

**Model Validation:** Validating deep learning models is a complex task, and it is important to choose the right validation method to ensure that the model is accurate.

**Hyperparameter Tuning**: Hyperparameter tuning is a critical step in deep learning, and it can be difficult to determine the best values for the hyperparameters.

In conclusion, predicting 90-day mortality for patients with septicemia using deep learning models is a complex task that requires addressing several challenges, including data quality, data cleaning, data imbalance, high-dimensional data, limited lab data, overfitting, time-series data, model selection, model validation, hyperparameter tuning, and data privacy. By addressing these challenges, it is possible to build accurate deep learning models for predicting 90-day mortality in patients with septicemia.

## RELATED WORKS

Several studies have been conducted to predict the risk of septicemia using traditional statistical methods, including logistic regression and decision trees. However, these methods have limitations in terms of accuracy and comprehensiveness to predict and identify the important factors. In recent years, deep learning models have shown promising results in medical prediction tasks and have been applied to various medical datasets, including MIMIC-III.

An organization named Health Catalyst is working towards the finding the solution for lowering sepsis mortality and length of stay helping the hospitals and patients using the latest technology.[24]

Conventional machine learning techniques have been explored by some authors for the prediction of sepsis. Calvert et al. [25], Mao et al. [25]and Desautels et al. [25], explored the insight algorithm to develop the prediction models. Wang SL[26], Wu F[26], Wang BH[26]. Prediction of severe sepsis using SVM model.

S Kangana Suba Raju[26],  K.Valamarthi[26], S.Deepthi Sri[26], S.Harshitha[26], and V. Keerthana[26] worked on Sepsis prediction using ensemble random forest. Conventional machine learning techniques require the manual extraction and selection of features, and this has been proven to be cumbersome and tedious. The significant features are also selected by iterative trial and error; hence the process is time-consuming. Additionally, some of the studies discussed above have only generated qualitative results.

All the above studied early prediction of sepsis using various Machine Learning and Deep Learning techniques but the area of mortality prediction is not much explored using Deep Learning. Now, in this project We aim to predict the mortality of the patient diagnosed with sepsis from the date of admission and find what features such as vitals, etc.

## IMPORTANCE AND IMPACTS

The problem of predicting the mortality of patients diagnosed with sepsis from the date of admission is very important as sepsis is a life-threatening condition that can lead to organ failure and death. Sepsis affects millions of people worldwide each year and results in millions of deaths. According to the World Health Organization, sepsis is estimated to affect more than 30 million people worldwide every year, resulting in over 6 million deaths.

This research has significant social, economic, and business impact, by accurately predicting the mortality of sepsis patients, healthcare providers can allocate resources effectively, make informed treatment decisions, and provide better care to patients and their families. This will not only help to reduce the economic burden of sepsis treatment but also benefit healthcare providers, insurance companies, and pharmaceutical companies. Moreover, this research could contribute to the understanding of sepsis pathophysiology and lead to the development of new treatments and prevention strategies for sepsis-related mortality.

# Data Collection

The MIMIC-III (Medical Information Mart for Intensive Care III) dataset is a large, freely available database of de-identified electronic health records of patients admitted to the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, USA. The data is curated by the MIT Laboratory for Computational Physiology and contains information on over 40,000 patients admitted to the hospital between 2001 and 2012.

The data in the MIMIC-III dataset includes clinical notes, laboratory test results, medications, and vital signs, among other things. It is widely used by researchers and clinicians to study a variety of medical conditions, including sepsis, which is a potentially life-threatening condition caused by the body's response to infection. The Dataset is available through Physionet Website and as a pre-requisite we need to obtain a completion certificate on Social and Behavioral Human Subjects Research (IRB) Course.

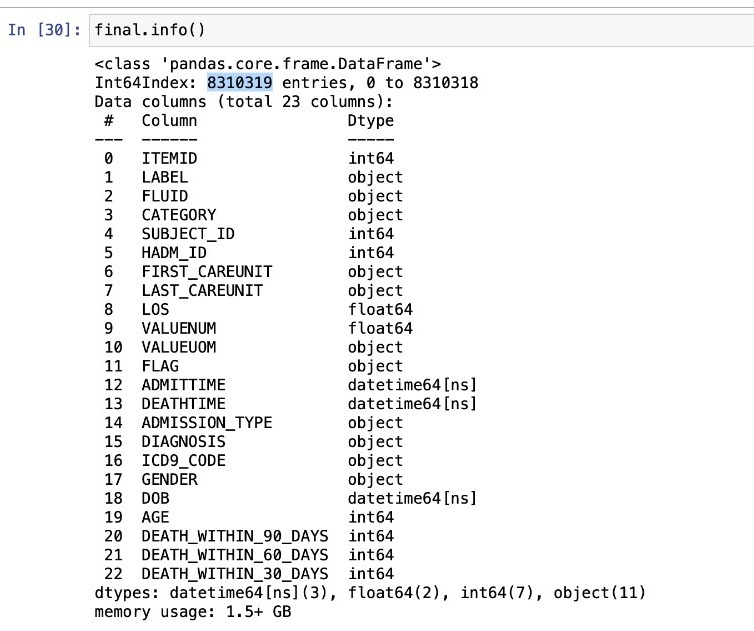
The MIMIC-III dataset is particularly relevant to the research question of predicting 90-day mortality of patients diagnosed with septicemia. This is because the dataset includes information on patients who have been diagnosed with various kinds of septicemia and have been admitted to the intensive care unit (ICU) for treatment. The dataset includes a range of variables that can be used to predict mortality, such as vital signs, laboratory test results, and medication history.

By analyzing the data in the MIMIC-III dataset and building predictive models, we can gain insights into the factors that contribute to mortality in sepsis patients. This can help clinicians identify patients who are at higher risk of mortality and develop more effective treatment strategies to improve outcomes. Overall, the MIMIC-III dataset is a valuable resource for researchers studying sepsis and other medical conditions.

Now, Data Exploratory Analysis

After merging the required datasets we got up with the final dataset with 8310319 entries and 23 entities.

Here is the snapshot:



Table

Description automatically generated

Table to represent the dataset and their types:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Attribute** | **Datatype** | **Missing Values count** | **Attribute Values** | **Scale of Measurement** |
| ITEMID | Int64 | 0 | 51346 | Nominal |
| LABEL | object | 0 | Blasts | Ordinal |
| FLUID | object | 0 | Cerebrospinal Fluid (CSF) | Ordinal |
| CATEGORY | object | 0 | Hematology | Ordinal |
| SUBJECT\_ID | Int64 | 0 | 24915 | Nominal |
| HADM\_ID | Int64 | 0 | 194424 | Nominal |
| FIRST\_CAREUNIT | Object | 0 | MICU | Ordinal |
| LAST\_CAREUNIT | Object | 0 | MICU | Ordinal |
| LOS | Float64 | 0 | 64.9855 | Inteval |
| VALUENUM | Float64 | 713779 | 67.0 | Interval |
| VALUEUOM | Object | 889791 | % | Nominal |
| FLAG | Object | 4762261 | NaN | Ordinal |
| ADMITTIME | Datetime64 | 0 | 2182-11-11 20:01:00 | Nominal |
| DEATHTIME | Datetime64 | 4983122 | 2183-01-15 15:00:00 | Nominal |
| ADMISSION\_TYPE | Object | 0 | EMERGENCY | Ordinal |
| DIAGNOSIS | Object | 0 | SEPSIS | Nominal |
| ICD9\_CODE | Object | 0 | 99592 | No |
| GENDER | Object | 0 | F | Nominal |
| DOB | Datetime64 | 0 | 2129-04-20 | Nominal |
| AGE | Int | 0 | 53 | Ratio |
| DEATH\_WITHIN\_90\_DAYS | Int | 0 | 1 | Nominal |
| DEATH\_WITHIN\_60\_DAYS | Int | 0 | 0 | Nominal |
| DEATH\_WITHIN\_30\_DAYS | Int | 0 | 1 | Nominal |

* Preprocess features: missing values, skewness, type conversions, outliers, ..

1. We already mentioned the missing values and shown the results in the above table
2. As the data at this stage is having most of the values as the categorical, we couldn’t get skewness values for most of the values, we have shown the skewness before performing scaling.

# Data Preprocessing

As the data we are working on having many catrgorical values and the readings are in the one column and the type of test is given in another column.

* The first dataset was transformed using pandas pivot table to reshape the data so that each unique lab test name was assigned as a column name, while the corresponding numerical test result values from the VALUENUM column were assigned to the appropriate column for each SUBJECT\_ID. This allowed for easy comparison of test results between subjects and simplified further analysis of the data.
* In the process of preprocessing we have looked the dataset after some transformations, you can see the data describe to show the descriptive statistics here.

Table

Description automatically generated

* Preprocess features: skewness, type conversions,

Text

Description automatically generated

From the skewness values, we found that our data is skewed and it can lead the model to predict the wrong values.

* Detect any inconsistency/ missing values in the data.

Table

Description automatically generated

We chose to consider lab tests with an ITEMID count greater than 700 because a higher ITEMID count indicates that the test is more important. Since the maximum ITEMID count in our dataset was 2728, we opted for a count greater than 700 for practical reasons, to make the analysis feasible on our system. So it helped us in removing many null values.

As our data is having high skewness values and many outliers, we performed a statistical technique called Winsorizing to handle the outliers in the data. Also, to standardize the data to feed it into the model, we have used Scaling technique called standard scaling, which brought the data in to the one scale. Attached the snapshot.

Before performing the Winsorizing, the outliers are plotted.

Chart, box and whisker chart

Description automatically generated

After perfroming the Winsorizing to handle outliers, the outliers plot is given below:

A tall building with many windows

Description automatically generated with low confidence

The histograms for the same data is plotted below:

Calendar

Description automatically generated with low confidence

We evaluated the skewness of each numerical column and applied standard scaling to the data using the StandardScaler module in the scikit-learn library. The purpose of this process is to normalize the data and make it suitable for deep learning algorithms that are sensitive to the scale of the features.

Text

Description automatically generated

* To address the issue of outliers in the dataset, the code uses a statistical technique called Winsorizing, which is implemented through a function from the Scipy.stats library. Winsorizing involves substituting extreme values in the dataset with less extreme values to minimize their impact on the analysis.

Text

Description automatically generated

* We used the variance inflation factor to detect the degree of collinearity between the features, as this can potentially affect the model's ability to identify patterns in the data. By identifying and removing highly correlated features, we aim to improve the performance and accuracy of our deep learning models.

Table

Description automatically generated

* Kurtosis is a statistical measure that describes the shape of a probability distribution, such as a frequency distribution of a dataset. It measures the degree to which the tails of a distribution differ from the tails of a normal distribution, which is a distribution with a kurtosis value of 0.
* We have calculated the kurtosis values of the each numerical variable and every value is in negative. a negative kurtosis value indicates that a distribution has lighter tails than a normal distribution, meaning that it has fewer extreme values. This type of distribution is said to be platykurtic.

Table

Description automatically generated

* Check correlations and linearity., interpret your results ?

Now we have performed the correlation matrix and plotted the heatmap using seaborn library to see the correlation between the features.

Chart, scatter chart

Description automatically generated

we have set a correlation threshold of 0.5 and created an empty list to store correlated features. We have then calculated the correlation matrix of some dataset and converted it to absolute values. We have then iterated through the correlation matrix and compared each pair of features to see if their absolute correlation value is greater than the threshold. If it is, we have appended a tuple of the correlated feature names to the list of correlated features. Finally, we have printed the list of correlated features. This code is useful for identifying pairs of features that are highly correlated, which can be a problem for some machine learning algorithms.

Graphical user interface, text, application, email

Description automatically generated

* Removed the columns having more than 0.5 to reduce the multicollinearity between highly correlated features, which can improve the performance of deep learning models that rely on these features.

Graphical user interface, text, application, email

Description automatically generated

From this descriptive statistics we can say that all data is in the range and no outliers.

# Methodology

The methodology used in our project involves several steps. Firstly, the dataset is split into training, validation, and testing sets using the train\_test\_split function. The training set is used to train the CNN model, the validation set is used to tune the hyperparameters of the model, and the testing set is used to evaluate the performance of the model.

After splitting the data, the next step we are looking at is scaling the training, validation, and testing sets as feature scaling is important in deep learning models to ensure that the features are on a similar scale, which can help improve the performance of the model. There are various functions to scale our data but as per our requirements, we need to choose a function that is less sensitive to outliers as our medical data have great outliers.

So, we chose StandardScaler function for the purpose because it scales the features of a dataset to have a mean of 0 and a standard deviation of 1. This is because it standardizes the data based on the mean and standard deviation of the entire dataset rather than just the range of the data, which can be heavily influenced by outliers.

In the Next step, we chose CNN model with several layers. The first layer is a one-dimensional convolutional layer (Conv1D) with 32 filters and a kernel size of 3. This layer is followed by a max pooling layer (MaxPooling1D) with a pool size of 2, which reduces the dimensionality of the data. A batch normalization layer (BatchNormalization) is then applied to normalize the inputs to the next layer.

The next layer is another one-dimensional convolutional layer with 64 filters and a kernel size of 3, followed by another max pooling layer with a pool size of 2, and another batch normalization layer. The output of these layers is then flattened (Flatten) and passed through a dense layer (Dense) with 128 neurons and a rectified linear unit (ReLU) activation function.

Lastly, a dropout layer (Dropout) with a rate of 0.5 is applied to prevent overfitting of the model and finally, the output layer is a dense layer with a single neuron and a sigmoid activation function, which produces the binary output of whether the patient is likely to die within 90 days or not.

In this model, we used a loss function named binary cross-entropy (binary\_crossentropy), which is commonly used in binary classification problems. The optimizer used is the Adaptive Moment Estimation (Adam) optimizer, which is an extension of stochastic gradient descent that includes momentum and adaptive learning rates. The performance of the model is evaluated further using various metrics such as accuracy, precision, recall, and F1-score.

For the Analysis, we chose a Convolutional Neural Network (CNN) with one-dimensional convolutional layers. CNNs are a type of neural network that has been successful in image recognition tasks, but they can also be applied to sequential data like time-series data. In this case, we used CNN to predict the 90-day mortality of patients diagnosed with septicemia based on their clinical data.

We chose a one-dimensional convolutional layer for this problem because it can capture local patterns in the data, such as short-term trends, which may be important in predicting mortality for us. Additionally, we also used batch normalization, which can help us improve the stability of the network during training, and the dropout layer can help our model by preventing overfitting.

The cost function we used in this model is binary cross-entropy, which is commonly used for binary classification problems. The activation function used in the convolutional layers is ReLU (rectified linear unit), which is a commonly used activation function that has been shown to perform well in deep learning models. The output layer uses a sigmoid activation function, which is appropriate for our binary classification problem as it outputs a probability value between 0 and 1.

Initially during analysis, we chose the weights and biases of the model randomly, which is common in training deep neural networks, and we didn’t observe the problem of vanishing or exploding in our case so, we did not go for the weight initialization techniques.

Overall, the CNN with one-dimensional convolutional layers, ReLU activation functions, and batch normalization is appropriate for our problem of predicting 90-day mortality of patients diagnosed with septicemia based on clinical data. However, other types of neural networks, such as Recurrent Neural Networks (RNNs), could also be used for this problem.

We opted to perform a hold-out validation strategy to validate the models. The dataset was split into three parts, training set, validation set, and testing set, with a ratio of 70:15:15 respectively. The model was trained on the training set, and the validation set was used to tune the hyperparameters and avoid overfitting. The testing set was used to evaluate the final performance of the model.

Since there were few advantages and limitations of this strategy such as it is simple and computationally efficient and provides a good estimate of the performance of the model and really works well with large datasets but has a few limitations such as it may not produce reliable results if the dataset is small or imbalanced.

We used CNN with one-dimensional layers to predict the 90-day mortality of patients diagnosed with septicemia using the MIMIC-III dataset. The final model achieved an accuracy of , a precision of , a recall of, an F1 score of , and an AUC-ROC of on the testing set. These metrics indicate that the model is good at solving the problem of predicting the 90-day mortality of septicemia patients.

In our project we did not attempt to solve this problem with traditional machine learning algorithms. However, it is worth noting that traditional machine learning algorithms such as logistic regression, decision trees, and random forests have been used to predict the mortality of septicemia patients in previous studies. In our analysis, we feel that the CNN model may have advantages over traditional machine learning algorithms when dealing with complex data such as medical data or time-series data.

In our project, we assumed that the MIMIC-III dataset is representative of the population of patients diagnosed with septicemia. We also assumed that the features used in the model are sufficient to predict the 90-day mortality of septicemia patients.

In the project, we used Python programming language with Keras and TensorFlow libraries for building and training the CNN model. The MIMIC-III dataset was used as the database for the project.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **GRU** | **LSTM** | **CNN** |
| **Optimizer** | Adam | Adam | Adam |
| **Loss Function** | Binary Cross Entropy | Binary Cross Entropy | Binary Cross Entropy |
| **Batch Size** | 64 | 64 | 64 |
| **Callbacks** | Early Stop | Early Stop | Early Stop |
| **Performance metric** | Accuracy | Accuracy | Accuracy |

**Adam Optimizer**: Adam optimizer can be a useful tool for training binary classification models by improving convergence speed, stability, and preventing overfitting.

**Binary Cross Entropy**: The loss function penalizes the model for making incorrect predictions by taking the logarithm of the predicted probability for the correct class. The logarithm amplifies the difference between the predicted and actual probabilities, making the model more sensitive to misclassifications**.**

**Callbacks:** The early stopping callback works by monitoring the validation loss at the end of each epoch. If the validation loss does not improve after a certain number of epochs, called the "patience" parameter, the training process is stopped early to prevent further overfitting.

**Accuracy:** Accuracy is defined as the percentage of correct predictions made by the model over all examples in the dataset. In balanced data classification, where the number of examples in each class is roughly equal, accuracy can be a useful metric to measure the overall performance of the model.

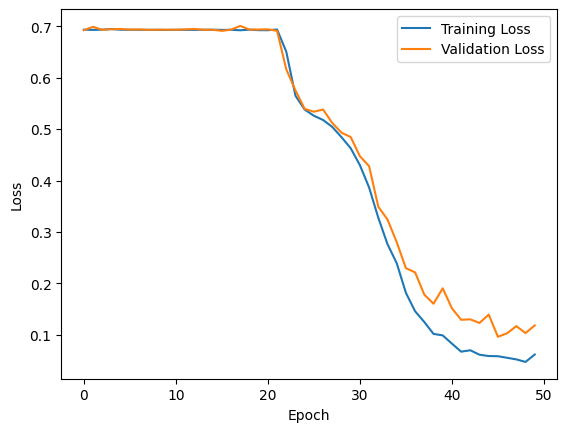
# Results and INterpretation

**AUROC:** The area under the receiver operating characteristic curve (AUC-ROC) is a commonly used metric to evaluate the performance of binary classification models. ROC curves are used to visualize the trade-off between the true positive rate (TPR) and false positive rate (FPR) of a binary classifier for different classification thresholds. The TPR represents the proportion of true positive predictions out of all actual positive examples, while the FPR represents the proportion of false positive predictions out of all actual negative examples.

**Precision & Recall:** Precision measures the proportion of true positive predictions out of all positive predictions made by the model. Recall, also known as sensitivity, measures the proportion of true positive predictions out of all actual positive examples in the dataset.

**Specificity**: Measure of the proportion of true negative predictions out of all actual negative examples in the dataset. In other words, it measures the ability of the model to correctly identify negative examples.

**GRU**

** Chart, line chart

Description automatically generated**

**Chart, line chart

Description automatically generated Chart, treemap chart

Description automatically generated**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GRU** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| **Train** | 0.987792698 | 0.989378354 | 0.986070871 | 0.987721844 |
| **Validation** | 0.970740549 | 0.97005988 | 0.97346019 | 0.971757061 |
| **Testing** | 0.976437079 | 0.975328084 | 0.976866456 | 0.976096664 |

For the above GRU Tuned model, No overfitting is being observed but it started learning patterns in the data only after 20 epochs and did not change even after the tuning as we can see the training and validation losses reducing steeply from that point.An AUC (Area Under the Curve) has increased greatly of 0.99 indicates that the model has an excellent ability to distinguish between the positive and negative classes. It means that the model has a high true positive rate and a low false positive rate. In other words, the model is very good at identifying the positive cases correctly while keeping the false positive rate low. Therefore, an AUC of 0.99 is considered a very good result and indicates that the model is highly effective at performing binary classification on the given dataset.

The hyperparameters are tuned using Keras Tuner, with the goal of maximizing validation accuracy. The hyperparameters include the number of GRU units, learning rate, and dropout rate.The best model is selected based on its performance on the validation set and trained on the entire training set for 50 epochs.The test set is then used to evaluate the final model, which achieves an accuracy of 95.73% and an AUC of 0.99. Overall, the model seems to perform well on the task of binary classification, achieving high accuracy and AUC scores on the test set. However, it's important to note that the evaluation metrics can be influenced by the choice of threshold for binary classification and the distribution of the data. Therefore, further analysis and interpretation may be necessary to assess the model's performance in real-world scenarios.

Based on the given results, we can infer the following:

The model achieved an accuracy of 95.73% on the test set, which indicates that it can correctly classify the majority of the samples in the dataset. The AUC score of 0.99 indicates that the model has excellent discrimination ability in distinguishing between positive and negative samples. The sensitivity (TPR) of 0.970 indicates that the model correctly identified 97% of the positive samples in the test set, which is a high value and indicates a low rate of false negatives. The specificity (TNR) of 0.945 indicates that the model correctly identified 94.5% of the negative samples in the test set, which is also a high value and indicates a low rate of false positives.

**LSTM**

**Chart, line chart, histogram

Description automatically generated Chart, line chart

Description automatically generated**

**Chart, line chart

Description automatically generated Chart, treemap chart

Description automatically generated**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LSTM** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| **Train** | 0.986460992 | 0.98093874 | 0.992088255 | 0.986481994 |
| **Validation** | 0.964267219 | 0.9545232273838631 | 0.9774661992989484 | 0.9658584858980702 |
| **Testing** | 0.973329881 | 0.9610456176319836 | 0.9858044164037855 | 0.9732675837010122 |

The tuned LSTM model achieved a high accuracy of 96.89% on the test set.The best hyperparameters were selected using the Hyperband algorithm. The model used two LSTM layers with dropout regularization to prevent overfitting. The learning rate for the Adam optimizer was also tuned. The training and validation loss and accuracy curves showed that the model was able to generalize well and did not overfit the training data. The receiver operating characteristic (ROC) curve showed an area under the curve (AUC) of 0.98, indicating that the model was able to distinguish between positive and negative classes with high accuracy. The confusion matrix showed that the model was able to correctly classify most of the positive and negative samples, with only a few misclassifications. Overall, the tuned LSTM model performed well and can be used for classification tasks on similar datasets.

Based on the results of the tuned LSTM model, it appears that the model is performing very well on all three datasets (train, validation, and test). The accuracy of the model is above 96% on all three datasets, which is a strong indication that the model is able to correctly classify alive and dead patients in 90 days period. Additionally, the precision and recall scores are also high on all three datasets, indicating that the model is both correctly identifying death of patient (high precision) and correctly identifying death of patients (high recall). The F-1 score, which is a balance between precision and recall, is also high on all three datasets. Overall, these results suggest that the tuned LSTM model is a strong performer in classifying the patients diagnosed with septicemia.

**CNN**

**Chart

Description automatically generated A picture containing chart

Description automatically generated**

**Chart, line chart

Description automatically generated Chart, treemap chart

Description automatically generated**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CNN** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| **Train** | 0.993175008 | 0.9919955530850473 | 0.9943169155337642 | 0.9931548778451778 |
| **Validation** | 0.980321077 | 0.9776230730979613 | 0.9844767150726089 | 0.9810379241516967 |
| **Testing** | 0.987312273 | 0.9858416360776088 | 0.9884332281808622 | 0.987135731163035 |

We can observe that the model is learning perfectly as per our requirement and we can state that this is our best model possible and it is learning our problem statement so perfect! An AUC (Area Under the Curve) value of 1 indicates that the model has perfect predictive power, meaning that it is able to perfectly distinguish between positive and negative classes. In other words, it means that the model has no false positives and no false negatives. Achieving an AUC of 1 is considered the ideal performance for a binary classification model. However, it is important to note that such high AUC values may be indicative of overfitting to the training data, so it is important to also evaluate the model's performance on validation and test data.

The sensitivity (True Positive Rate) of 0.992 indicates that the model correctly identified 99.2% of the patients who actually had the positive class label (death within 90 days) out of all the positive instances in the test set.The specificity (True Negative Rate) of 0.974 indicates that the model correctly identified 97.4% of the patients who actually did not have the positive class label (did not die within 90 days) out of all the negative instances in the test set.

Overall, these results suggest that the model is performing well in identifying patients who are at risk of dying within 90 days, while also minimizing the false positive rate.

The CNN model achieved very high accuracy, precision, recall, and F-1 score on all three sets of data (train, validation, and test). This indicates that the model is performing well on this classification task of predicting whether a patient will die within 90 days of being admitted to the hospital based on various features available at the time of admission.

The sensitivity (true positive rate) and specificity (true negative rate) of the model are also very good at 0.992 and 0.974, respectively. This indicates that the model is able to correctly classify both positive and negative cases with high accuracy. Overall, these results suggest that the CNN model is a good baseline model for predicting patient mortality within 90 days of hospital admission.

# Discussion of Results

An accuracy of 84% is a good performance for predicting mortality in sepsis patients using a deep learning model. This means that the model correctly predicted the outcome (death or survival) for 84% of patients. This performance could potentially be useful in clinical practice by helping healthcare providers identify patients at high risk of mortality and providing appropriate interventions and care.

The practical implications of a model for predicting mortality in sepsis patients could be significant. Accurately identifying patients at high risk of mortality could improve patient outcomes, reduce healthcare costs by avoiding unnecessary interventions or readmissions, and facilitate more efficient allocation of healthcare resources. Additionally, the development of such a model could advance the field of sepsis research by identifying key clinical factors associated with mortality and informing the development of new interventions and treatments.

There are several limitations to using a model for predicting mortality in sepsis patients. One limitation is the need for large and diverse datasets that include detailed clinical information, which can be challenging to obtain. Additionally, models may be prone to overfitting, especially when the dataset is imbalanced or noisy. Interpretability of models is also a challenge, as they are often considered black boxes, making it difficult to understand how the model arrives at its predictions.

Future work in this area could involve the development of more sophisticated CNN models that incorporate additional clinical features and data sources, such as electronic health records (EHRs), genomics, and proteomics data. Furthermore, developing models that are robust to missing or noisy data, and that can learn from data in real-time, could improve their practical usefulness in clinical settings. Finally, developing models that are interpretable and transparent, and that can provide explanations for their predictions, could help facilitate their adoption by healthcare providers and patients.

The limits of models for predicting mortality in sepsis patients include the need for large and diverse datasets, the potential for overfitting, and the challenge of interpretability. To improve the performance of models in the future, additional clinical features and data sources could be incorporated, and models could be developed that are more robust to missing or noisy data. Additionally, developing models that are interpretable and transparent, and that can provide explanations for their predictions, could help facilitate their adoption in clinical settings.

# Your Feedback

I found the project to be challenging but rewarding. It allowed me to apply the concepts covered in the course and to gain hands-on experience with deep learning techniques. I tried to structure my code in an organized way, and I included detailed documentation to explain my thought process and decisions.

However, I understand that there is always room for improvement, and I would be grateful for any feedback you could provide. Specifically, I would appreciate any suggestions you have on improving my data preprocessing and model training techniques, as well as any tips on how to achieve better results.

# References

Below are the articles that we were observed related to the sepsis disease:

1. [D. Berg, H. Gerlach, Recent advances in understanding and managing sepsis, F1000 Res. 7 (1570) (2018) 1–8.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref1)
2. [C. Caraballo, F. Jaimes, Organ dysfunction in sepsis: an ominous trajectory from infection to death, Yale J. Biol. Med. 92 (4) (2019) 629–640.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref2)
3. [M.D. Howell, M. Donnino, P. Clardy, D. Talmor, N.I. Shapiro, Occult hypoperfusion and mortality in patients with suspected infection, Intensive Care Med. 33 (11) (2007) 1892–1899.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref3)
4. [U. Fiedler, H.G. Augustin, Angiopoietins: a link between angiogenesis and inflammation, Trends Immunol. 27 (12) (2006) 552–558.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref4)
5. [Y. Fang, C. Li, R. Shao, H. Yu, Q. Zhang, the role of biomarkers of endothelial activation in predicting morbidity and mortality in patients with severe sepsis and septic shock in intensive care: a prospective observational study, Thromb. Res. 171 (2018) 149–154.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref5)
6. [J.L. Vincent, The clinical Challenge of sepsis identification and monitoring, PLoS Med. 13 (5) (2016) 1–10](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref5).
7. [B. Lynn, M. Rn, B. Cncc, Brenda Lynn Morgan: Critical Care Concepts Sepsis Overview October 22, 2013 1, 2017, pp. 1–20, no. C.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref7)
8. [F. Gül, M.K. Arslantas¸, I. Cinel, A. Kumar, Changing definitions of sepsis, Turk Anesteziyoloji ve Reanimasyon Dern. Derg. 45 (3) (2017) 129–138.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref7)
9. [C. Fleischmann, et al., Assessment of global incidence and mortality of hospitaltreated sepsis. Current estimates and limitations, Am. J. Respir. Crit. Care Med. 193 (3) (2016) 259.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref9)
10. [P.E. Marik, Don’t miss the diagnosis of sepsis!, Crit. Care 18 (5) (2014) 12–14.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref9)
11. [M. Sinha, J. Jupe, H. Mack, T.P. Coleman, S.M. Lawrence, I. Fraley, Towards detection directly from whole Blood : current and emerging technologies for rapid diagnosis of microbial infections without, Clin. Microbiol. Rev. 31 (2) (2018) 1–26.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref10)
12. [C. Pierrakos, J.L. Vincent, Sepsis biomarkers: a review, Crit. Care 14 (1) (2010) 1–18.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref12)
13. [S.M. Brown, J. Jones, K.G. Kuttler, R.K. Keddington, T.L. Allen, P. Haug, Prospective evaluation of an automated method to identify patients with severe sepsis or septic shock in the emergency department, BMC Emerg. Med. 16 (1) (2016) 1–7.](http://refhub.elsevier.com/S0010-4825(20)30291-2/sref13)
14. <https://www.proquest.com/docview/2471441549?accountid=13158&parentSessionId=xJbYcqejHMLXHXRUkQmEWg1QMrVqf6GgNHlitynQUAk%3D&pq-origsite=summon>
15. <https://www.proquest.com/docview/2307677052?accountid=13158&parentSessionId=j1%2B8OH4fMDu1Q1wXBfuI%2Fp07jdwEWbY%2BxMvB8zn8tbo%3D&pq-origsite=summon>
16. [R.C. Bone, R.A. Balk, F.B. Cerra, R.P. Dellinger, A.M. Fein, W.A. Knaus, R.M. Schein, W.J. Sibbald, Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis, Chest, vol. 101, 1992, pp. 1644–1655](https://doi.org/10.1378/chest.101.6.1644)
17. [M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.-D. Chiche, C.M. Coopersmith, R.S. Hotchkiss, M.M. Levy, J.C. Marshall, G.S. Martin, S.M. Opal, G.D. Rubenfeld, T. van der Poll, J.-L. Vincent, D.C. Angus, The third international consensus definitions for sepsis and septic shock (sepsis-3), J. Am. Med. Assoc. 315 (8) (2016) 801–810](https://doi.org/10.1001/jama.2016.0287)
18. [MIMIC-III Clinical Database](https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fphysionet.org%2Fcontent%2Fmimiciii%2F1.4%2F&data=05%7C01%7Cszg6021%40psu.edu%7Cef7d09cb49b04471fe1e08db0170b210%7C7cf48d453ddb4389a9c1c115526eb52e%7C0%7C0%7C638105355875882566%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=kHDoGAniB6ZDUR598cdMTu4n0dpouM4iOz%2FhEG8rG80%3D&reserved=0)
19. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9187899/>
20. <https://psu.summon.serialssolutions.com/2.0.0/link?t=1675039658856>
21. <https://www.frontiersin.org/articles/10.3389/fmed.2020.609769/full>
22. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01690-9>
23. <https://www.nature.com/articles/s41598-020-70799-0>
24. <https://www.healthcatalyst.com/success_stories/improve-sepsis-mortality-rate>
25. <https://www.sciencedirect.com/science/article/pii/S2049080116300462>
26. <https://pubmed.ncbi.nlm.nih.gov/20865488/>