2021 2nd Global Conference for Advancement in Technology (GCAT)
Bangalore, India. Oct 1-3, 2021

Early Sepsis Detection using Machine Learning and Neural Networks

Nemil Shah*
Student, Computer Engineering
DJ Sanghvi College of Engineering
Mumbai, India
shahnemil103@gmail.com

Jay Bhatia*
Student, Computer Engineering
DJ Sanghvi College of Engineering
Mumbai, India
jaybhatia73@gmail.com

Nimit Vasavat*
Student, Computer Engineering
DJ Sanghvi College of Engineering
Mumbai, India
vnimit1991@gmail.com

Rishi Desai*
Student, Computer Engineering
DJ Sanghvi College of Engineering
Mumbai, India
rishindesai17@gmail.com

Pankaj Sonawane*
Professor, Computer Engineering
DJ Sanghvi College of Engineering
Mumbai, India
pankaj.sonawane@djsce.ac.in

2021 2nd Global Conference for Advancement in Technology (GCAT) | 978-1-6654-1836-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GCAT52182.2021.9587778

***Abstract* — Sepsis can cause overwhelming changes that impair multiple organs, leading to their failure and sometimes even having fatal results. Low blood pressure, difficulty in breathing, fever, mental confusion, and fast heart rate are some of the symptoms. Infections acquired in health care settings also frequently result in Sepsis. These infections affect hundreds of millions of patients globally every year and are one of the most recurrent adverse events during care delivery. Clinical conditions can rapidly deteriorate as healthcare-associated disorders are often impervious to antibiotics. During critical scenarios such as an intensive care unit, Early diagnosis and timely medical treatment of sepsis is crucial to give the patient prompt treatment since there is an increase in the mortality rate as each hour passes away in critical care. This paper proposes a method to predict the inception of sepsis 6 hours in advance using various machine learning and deep learning models and presents a comparative study of the same. The Medical Information Mart for Intensive Care III (MIMIC3) dataset was used to test the traditional machine learning methods such as RandomForest(RF), XGBoost, and also Deep Learning techniques such as neural networks and Autoencoders with XGBoost. The dataset contains an extensive range of parameters that are associated with laboratory, vitals, and demographics of patients which help them classify as sepsis and non-sepsis patients. A data pipeline was successfully created to clean the data, impute missing values and perform various feature engineering techniques. Our best performing model is a Deep Neural Network with an AUC ROC score of 0.888.**

Keywords* — *Sepsis, Deep neural networks, Autoencoders, XGBoost, Random Forest, Machine Learning

I. INTRODUCTION

Sepsis can be defined as chronic critical illness related to severe inflammation, reduced immune and organ damage [1]. Sepsis is one of the major causes of death in the Intensive Care Unit (ICU) which affects more than 18 million all over the world with mortality rates as high as 30%. In 2013, the US health system accounted for $24 billion for the treatment of sepsis [2]. Prior detection and specific treatment has been identified as critical features to ameliorate the conditions of patients with sepsis. It can consequently reduce post-traumatic ramifications. A delay in the detection may increase mortality by approximately 4-8% [3]. Due to these reasons, sepsis has become a global public health issue that requires attention in developing ways for early detection of sepsis.

Currently, doctors identify sepsis by using the patient's vital signs and the symptoms, and individualistic biomarkers. But due to the intricacy of sepsis and various organ dysfunction, the biomarkers may differ for every patient [2]. Clinical Decision Rules (CDRs) are currently used for predictive analysis which lacks generalizability when applied to mass populations. Moreover, it may take years for CDRs to develop which makes it difficult for them to update when new information is accessible [4]. There is a requirement to enhance diagnosis and prediction which are more reliable and convenient than the traditional methods.

With the burgeoning of computerized data, machine learning (ML) models have become popular for predictive analysis [4]. A variety of ML models have been used for diagnosis which includes SVM, XGBoost, KNN, improvement in traditional Deep Forest [3], Deep Learning models [5], LSTMs [1], RNNs[4], and many more.

This paper proposes a method that can predict sepsis 6 hours earlier than the clinical prediction. This paper also provides a comparative study of different ML models which are trained and trained on a publicly available dataset. An ML predictive model allows us to identify relations between different variables and discover patterns that are otherwise inconspicuous in traditional predictive methods.

II. LITERATURE SURVEY

The following papers focused on predicting sepsis using Machine Learning and providing results. Md. Mohaimenul Islam et al [6] performs a literature survey of sepsis papers published between 2000-2018 and concludes that the machine learning approach to predict sepsis had a better performance than the traditional sepsis scoring systems such as SIRS, MEWS, and SOFA. Desautels et al [7] have applied the InSight machine learning classification system to predict sepsis using multiple variables and compared its results with systemic inflammatory response syndrome (SIRS), modified early warning score (MEWS), simplified acute physiology score (SAPS) II, and quick sequential organ failure assessment (qSOFA). Their results show that their machine learning model InSight produced superior performance on detecting sepsis onset in comparison with the other methods. They have even shown that the model also performs well in the case of randomly missing data.

M. Saqib et al [2] have compared the results of the traditional machine learning methods, LSTM and Attentional

'*' indicates equal contribution

LSTM for early Sepsis Prediction, and a comparative study is documented. Various statistical techniques such as the chi-squared test of independence were used to determine the best features which were to be used as input data. Some feature engineering techniques were used to determine how some features are directly and inversely related to the Sepsis Label. The proposed method states that RF was the most successful model and LSTM models trained with attentional mechanisms only achieved a slight but noticeable improvement over LSTM models trained without attentional mechanisms.

S. D. Wickramaratne et al [4] have proposed a GRU Network as a model for Sepsis Detection beforehand. All vital, laboratory and demographic data are used by the proposed model to predict sepsis. Their analysis suggests that GRUs make each recurrent unit capture dependencies of different time scales adaptively. The proposed model uses Bi-Directional GRU because over LSTM, due to a lesser number of parameters the GRU model uses less computational time and converges faster, thus providing a better result than LSTM in some cases. The proposed model used two classifiers, one which included only the vital signs as features and the other included both vital signs and laboratory values, the latter giving better accuracy. Traditional baseline models such as Logistic Regression(LR), Random Forests(RF), and Support Vector Machines(SVM) gave a poor performance than The proposed GRU model with regards to sensitivity and specificity.

Bedoya et al [8] have trained a deep learning model which is a multi-output Gaussian process and Recurrent Neural Network(MGP-RNN). Their training set and internal validation included 42 979 encounters, while their temporal validation set included 39 786 encounters. They made comparisons to Random forest, cox regression, penalized linear regression and the clinical scores used to detect sepsis were SIRS, qSOFA, and NEWS. The C-statistic of their model predicting sepsis within 4 hours of onset was higher than the other models. Their model detected sepsis 5 hours early. The temporal validation assessment continued to show higher performance. L. Tran [9] et al has proposed a deep neural network called AEC-Net which concurrently optimizes an auto-encoder and a fully connected neural network. Another model is also proposed which is an ensemble of AEC-Net, RF and GBDT. Both of these achieved better performance than the baseline models(RF and XGB) taken by the authors.

M. Fu et al [3] have used the MIMIC-III dataset consisting of a total of 3125 patients for early prediction of sepsis and do a comparative study of different machine learning models. Their improved cascade deep forest model includes two Random Forest and two XGBoost in each layer which performs better than the traditional machine learning models. It makes use of k-fold cross-validation.

S. Sarafrazi et al [12] have provided a comparative study of different models to predict sepsis 12 hours before its diagnosis. Various sequential models such as CNN, LSTM, CNN_LSTM, and XGBoost model are trained with XGBoost outperforming all the other models including XGBoost-CNN-LSTM. Feature Engineering is also performed which results in a 5% improvement in the XGBoost model. G. Tsang et al [5] have proposed a fresh deep learning model for sepsis detection up to 6 hours in advance. The novel model is a combination of cascade and boosting algorithms that improve upon the imbalance of traditional methodologies and provides a more balanced method. The hinge loss function prevents the model from over-fitting over the increasing parameter-space complexity of boosted cascade methods.

III. System Methodology

A. System Workflow

The MIMIC-3 Dataset was used for our proposed method. The dataset was first grouped patient-wise for better analysis and further prediction. Huge datasets obtained from public repositories usually contain redundant data which has to be cleaned efficiently. The missing values were imputed patient-wise. Various methods were analyzed for filling in the missing data and the most efficient method was considered. Various data preprocessing steps were used to undersample the data because of a huge imbalance between the positive and negative values. Feature Selection was performed wherein the most important features were considered as input data for the models. Different models were trained and hyperparameter tuning was performed. A comparative study of their results is provided in our proposed method.
Refer Fig 1 to navigate through the system workflow.

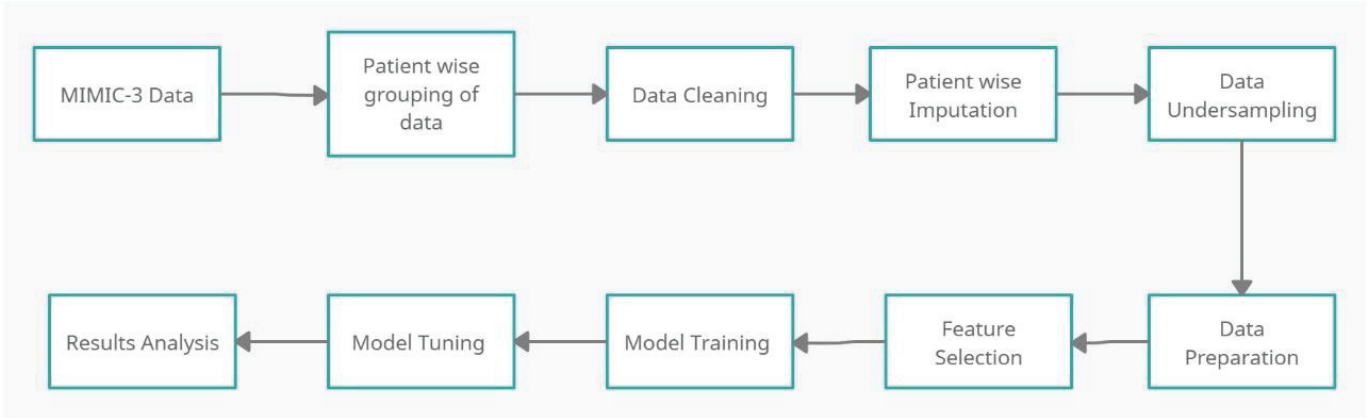

Fig. 1. System Workflow

B. Dataset Description

The dataset used is from PhysioNet Computing in Cardiology Challenge 2019 [11] which consists of 40,336 records of patients with 2932 patients detected positive with sepsis and 37,404 patients not detected with sepsis.

This dataset is collected from two different hospitals providing patient-level hourly based records with 40 time-dependent variables divided into three categories - vital signs, lab tests results, and demographic information. The vital signs values were accessible for majority of the records averaging about 32.4% missing data but the majority of the values were missing from the lab results averaging 94.9%. As for the demographic features, there is no data missing.

Target labels are provided for hourly timestamps as either 0 or 1, negative or positive.

The dataset consists of 22,566 males and 17,770 females with an average age of 61.6 years. The minimum and maximum age recorded is 14 and 90+ respectively.

C. Data Preprocessing

The MIMIC3 dataset was first extracted and all the individual patient files were combined in comma-separated values (CSV) format. Each row was given a patient id so that the same patient rows can be grouped. The laboratory values were recorded periodically and hence most of them had missing percentages of more than 90%. Fig 2 shows the percentage of missing data in each feature.

Heart rate (HR) and Pulse oximetry (O2Sat) are important features in predicting Sepsis. The patients with more than 50% null values for these two features were removed. Patients with overall missing data greater than 65% were also removed. Since our goal was to predict sepsis for adults, patients with age less than 18 were also removed.

Initially, the remaining missing data were replaced with mean and median for each patient but the results came out better when the missing data was imputed with forwarding fill and backward fill techniques for each patient.

Forward fill propagates the last observed non-value forward until another non-null value is encountered. Backward fill propagates the first observed non-null value backward until another non-null value is met. Forward fill was applied to missing values in tail rows of a patient and backward fill was applied to missing values in head rows of a patient. After all the above steps, the columns which still had a majority of their values missing throughout the dataset were removed. Four such columns were found namely, End-tidal CO2 (EtCO2), Fibrinogen, Bilirubin direct, and Troponin I (TroponinI). Physiologically normal values were considered for the rest of the features whose entire column was missing for a patient.

Only 24630 patient records were left after preprocessing. But this data is still unbalanced. Among the 24630 patients left, 21698 were negative for sepsis and 2932 were positive for sepsis. If this dataset is trained, our model would be biased. Hence, an undersampling technique was applied to balance the dataset. The sepsis negative patients were randomly deleted so that they can match with the sepsis positive patients. The final dataset contained 5928 patients, out of which 2998 were negative for sepsis and 2930 were positive for sepsis.

Instead of splitting the data by 80-20 split, patient-wise splitting was done to prevent data leakage. In an 80-20 split, the rows of the dataset are split randomly and hence some patients might be present in training as well as testing set. This causes the model to be overfitted. Hence, a patient-wise split is better where all the rows of a particular patient are either present in the training set or the testing set. 5000 patients were used for training and the remaining 928 patients were used for testing.

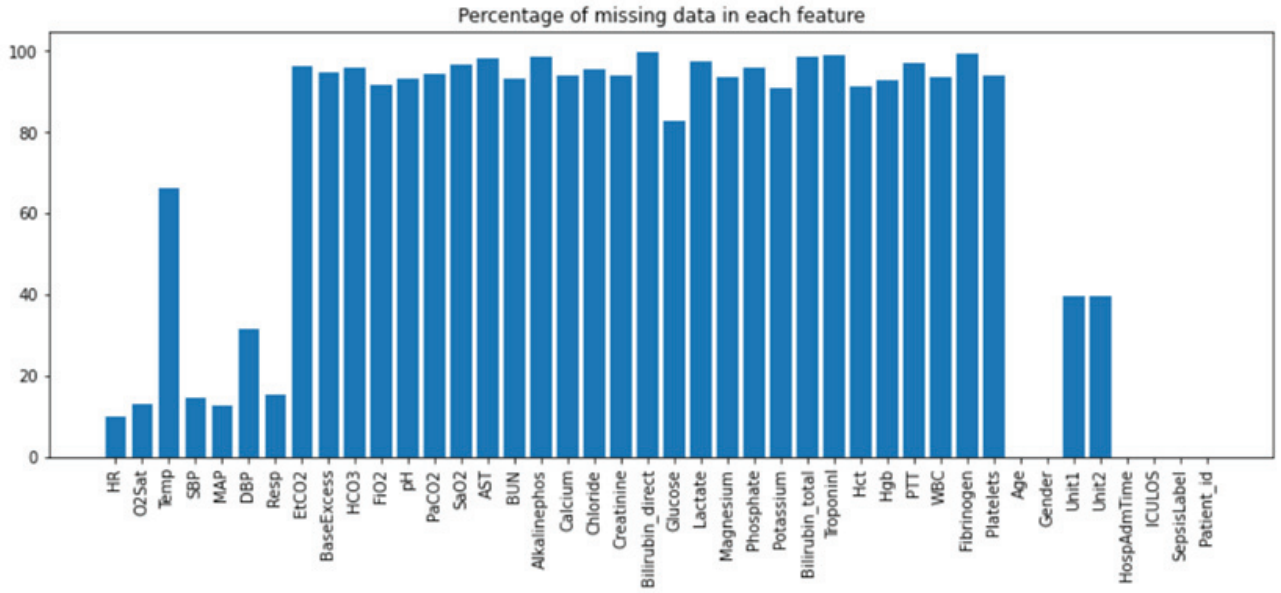

Fig. 2. Percentage of missing data in each column

IV. MODEL BUILDING

A. Traditional Machine Learning Techniques

Traditional machine learning models like XGBoost and Random Forest were trained. The test set was excluded in the calculation of F scores in feature importance to prevent data leakage. Top 37 features were included in training both the XGBoost and Random Forest models. Table I includes the F scores of the top 14 features. Hyperparameter tuning was done in both models.

For XGBoost, the hyperparameters mostly included booster parameters such as maximum depth, minimum child weight, gamma, subsample, colsample by a tree, and regularization parameters. Learning parameters i.e learning rate was also tuned. Even though the number of positive and negative patients of sepsis were almost equal, the hourly cases for sepsis labels of all patients were still unbalanced. Hence, the booster parameter scale_pos_weight was also used to improve results. The value of this hyperparameter was kept as the ratio of negative sepsis labels to positive sepsis labels.

TABLE I. F SCORE OF TOP 14 FEATURES

Features	**F score**	**Features**	**F score**
ICU length of stay (ICULOS)	88	Creatinine	25
White Blood Cells (WBC)	51	Aspartate transaminase (AST)	24
Temperature (Temp)	47	Oxygen saturation from arterial blood (SaO2)	24
Hours between hospital admit and ICU admit (HospAdmTime)	42	Partial thromboplastin time (PTT)	23
Platelets	29	Blood urea nitrogen (BUN)	22
Calcium	28	Alkaline phosphatase (Alkalinephos)	17
Fraction of inspired oxygen (FiO2)	28	Bilirubin direct	16

For Random Forest, hyperparameters that were tuned included n estimators, minimum samples leaf, split criteria, bootstrap, the maximum number of features kept, maximum depth, and n jobs. Similar to scale_pos_weight, class weight hyperparameter was used in Random Forest to tackle the sepsis label imbalance. To further tackle the sepsis label imbalance, a Balanced Random Forest Classifier was used which randomly undersamples each bootstrap sample to balance it. Better results were achieved by using the Balanced Random Forest Classifier.

B. Autoencoders with XGBoost

Autoencoders can be classified as a neural network method that is trained to study a compressed representation of raw data. In the proposed model autoencoder is used as a classification predictive model for the Sepsis Label wherein, the compressed representation of the input features is learned by the model. The design of the autoencoder is such that the input data is restricted to the bottleneck up to the midpoint of the model and then the data is reconstructed generating pseudo input data.

It is an unsupervised approach for learning a lower-dimensional feature representation from training data. Autoencoder consists of encoder and decoder layers. Encoder learns the mapping from data to a low-dimensional latent space where data can be compressed into a small latent vector and compact and rich feature representation can be learned. Decoder learns to map back from latent space to reconstructed training data. Before training the Autoencoder the training data is normalized using a MinMaxScaler. An autoencoder model has been created in which only the non-Sepsis cases. The model will try to learn the best representation of non-Sepsis cases. This is because the autoencoder will try to learn only one class and automatically distinguish the other class.

The reconstruction loss should be as minimal as possible as it forces the latent representation to encode the majority of information about the data as possible into a lower-dimensional latent space while still being able to generate correct representation. Another model consisting of sequential layers is constructed and the trained weights till the layer of the latent representation of the input trained model in the architecture are added in this model. The hidden representation of the sepsis and non-sepsis labels are thus generated by predicting the input data using the above sequential model. The new reconstructed input data generated by the autoencoder is used as training data for the proposed XGBoost Model defined previously and the results of the two models are compared.

C. Deep Neural Network

In this section, we present our proposed Deep Neural Network(DNN) model. DNN is a good choice not only due to its property of approximating functions but also due to its feature learning capacity [9]. It is a powerful ML method with a major drawback, overfitting, which we have controlled using early stopping. The model contains three hidden layers with 128, 128, 128 cells respectively, and rectified linear activation or Relu is used to transform summed weighted input from a hidden layer to input to another layer. The Relu function and its derivative both are monotonic. It returns 0 if the input is less than zero i.e. negative and returns 1 if the input is equal to or greater than zero i.e. positive. Hence it ranges from 0 to infinity.

Adam optimizer is used as an optimization technique for gradient descent. The method is effective when the data is large and involves a lot of parameters. It uses an average of second moments of gradients instead of only adapting the learning parameter rates based on the average first-rate.

$$w_{t+1} = w_t - \alpha m_t$$

$$m_t = \beta m_{t-1} + (1-\beta)\left[\frac{\delta L}{\delta w_t}\right]$$

Eq. (1)

In Eq. (1) w represents weight at a given time, m is the aggregate of gradients at a given time, β is the moving average parameter, is the learning parameter and δL represents derivative of a loss function.

V. Results and Discussion

The feature importance was calculated with the help of SHAP values from the SHAP library. The SHAP values can collectively describe how each feature is correlated, either positively or negatively to the target variable. It shows the positive or negative relation of each feature with the target variable. The global interpretability of the features is displayed in Fig 3 and Fig 4 using SHAP values.

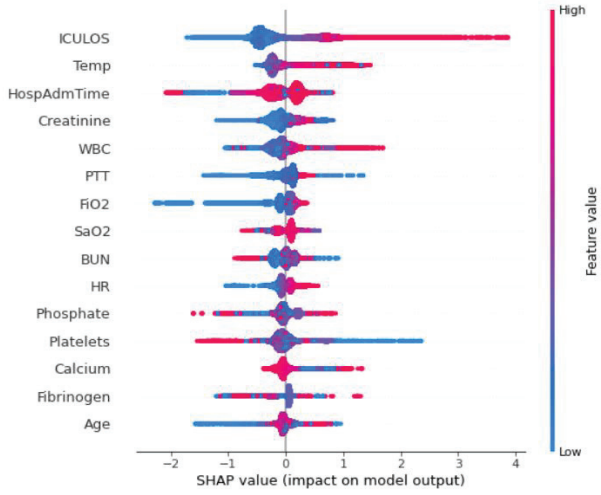

Fig. 3. SHAP Value of top 15 Features

In Fig 3 the variables are ranked in decreasing order of their feature importance. The horizontal position for each value in the features displays the association of the value with a higher or lower prediction and the color represents if the variable is high or low for the particular observation.

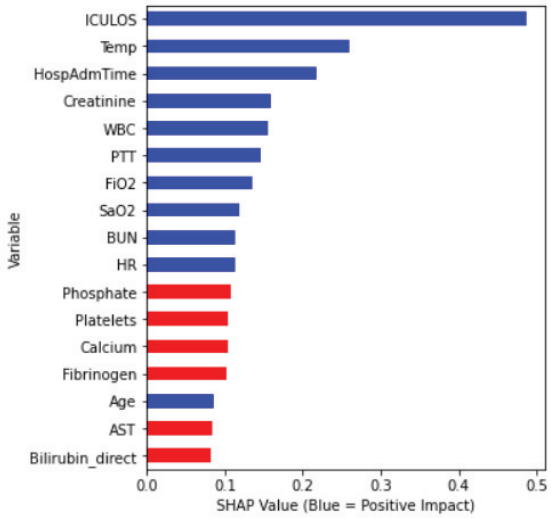

Fig. 4. Correlation Impact of Top Features

Fig. 4 shows the correlation effect of the top 15 features and if they are positively or negatively correlated. The features having a blue bar are positively correlated with the target variable and the features having a red bar are negatively correlated with the target variable. It is observed in Fig 3 that ICULOS has a high positive correlation with the target variable, whereas Phosphate has a high negative correlation with the target variable. It can also be inferred that features such as Temp, HospAdmTime, Creatinine, WBC, etc. have a positive correlation with the sepsis label of varying degrees. Other features such as Platelets, Calcium, and AST have a high negative correlation with sepsis labels.

The four models were evaluated based on five metrics - Accuracy, Sensitivity, Specificity, AUC ROC score, and F1 score.

As seen from Table II., the specificity for traditional machine learning techniques is good but the sensitivity is low and hence, these models cannot be completely trusted for positive label predictions. During the hyperparameter tuning of traditional machine learning techniques, it was observed that there was always a trade-off between accuracy and sensitivity. If the accuracy was increased, the sensitivity decreased, and if the sensitivity was increased then accuracy as well as specificity decreased. The model results shown were the best that could be tuned with a balance of sensitivity and specificity.

TABLE II. RESULTS

	Accuracy	**Sensi-tivity**	**Speci-ficity**	**AUC ROC**	**F1**
XGB	74.58	0.55	0.77	0.713	0.302
RF	76.12	0.51	0.79	0.712	0.301
Auto-XGB	84.31	0.58	0.87	0.816	0.417
DNN	81.08	0.86	0.77	0.888	0.819

By adding Autoencoders to the XGBoost model, the specificity of the model increases but there is not much effect on sensitivity. The accuracy of this model comes out to be the best among all four models.

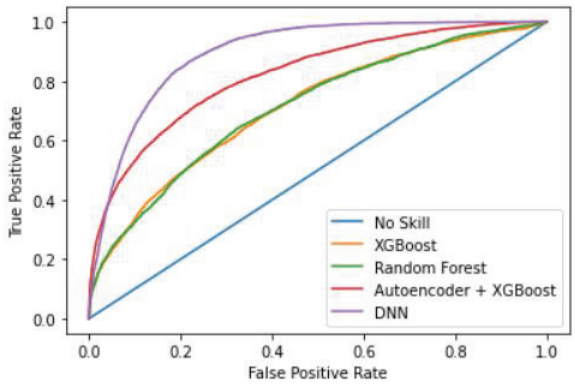

Fig. 5. AUCROC Curve

The Deep Neural Network model built has the highest sensitivity among all the models. Even though the accuracy and specificity of Autoencoder with XGB is higher than Deep Neural Network but the predicted positive label for

sepsis by the Deep Neural Network model will be more reliable due to its high sensitivity than other models. The AUC ROC score of the proposed models is displayed in a graphical format in Fig. 5.

In conclusion, Deep Neural Network is the best performing model out of all the proposed models.

VI. CONCLUSION

In our proposed system, the data preprocessing steps such as filling in missing data and filtering data were successfully carried out according to the model requirements, and clean pre-processed data was given as an input to our proposed system to predict sepsis. A Deep Neural Network (DNN) was proposed that beats the traditional machine learning techniques in terms of AUC ROC and F1 score. Autoencoder with XGBoost was also proposed but the DNN outperforms them too. Thus, the proposed system helps clinicians to predict sepsis 6 hours early.

In future work, more sepsis-positive patients can be added to the dataset for better accuracy and more precise models. Using hybrid models with advanced hyperparameter tuning can result in better accuracy of sepsis label prediction and can be more reliable.

REFERENCES

[1] T. Vicar, P. Novotna, J. Hejc, M. Ronzhina and R. Smisek, "Sepsis Detection in Sparse Clinical Data Using Long Short-Term Memory Network with Dice Loss," 2019 Computing in Cardiology (CinC), 2019, pp. Page 1-Page 4, doi: 10.23919/CinC49843.2019.9005786.

[2] M. Saqib, Y. Sha and M. D. Wang, "Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 4038-4041, doi: 10.1109/EMBC.2018.8513254.

[3] M. Fu, J. Yuan and C. Bei, "Early Sepsis Prediction in ICU Trauma Patients with Using An Improved Cascade Deep Forest Model," 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), 2019, pp. 634-637, doi: 10.1109/ICSESS47205.2019.9040774.

[4] S. D. Wickramaratne and M. Shaad Mahmud, "Bi-Directional Gated Recurrent Unit Based Ensemble Model for the Early Detection of Sepsis," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 70-73, doi: 10.1109/EMBC44109.2020.9175223.

[5] G. Tsang and X. Xie, "Deep Learning Based Sepsis Intervention: The Modelling and Prediction of Severe Sepsis Onset," 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 8671-8678, doi: 10.1109/ICPR48806.2021.9412058.

[6] Md. Mohaimenul Islam, Tahmina Nasrin, Bruno Andreas Walther, Chieh-Chen Wu, Hsuan-Chia Yang, Yu-Chuan Li, Prediction of sepsis patients using machine learning approach: A meta-analysis, Computer Methods and Programs in Biomedicine, Volume 170, 2019, Pages 1-9, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2018.12.027.

[7] Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman M, Barton C, Wales D, Das R Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach JMIR Med Inform 2016;4(3):e28

[8] URL: https://medinform.jmir.org/2016/3/e28

[9] DOI: 10.2196/medinform.5909

[10] Bedoya, Armando & Futoma, Joseph & Clement, Meredith & Corey, Kristin & Brajer, Nathan & Lin, Anthony & Simons, Morgan & Gao, Michael & Nichols, Marshall & Balu, Suresh & Heller, Katherine & Sendak, Mark & O'Brien, Cara. (2020). Machine learning for early detection of sepsis: an internal and temporal validation study. Journal of the American Medical Informatics Association Open. 3. 10.1093/jamiaopen/ooaa006.

[11] L. Tran, M. Nguyen and C. Shahabi, "Representation Learning for Early Sepsis Prediction," 2019 Computing in Cardiology (CinC), 2019, pp. 1-4, doi: 10.23919/CinC49843.2019.9005565.

[12] S. Sarafrazi et al., "Cracking the "Sepsis" Code: Assessing Time Series Nature of EHR Data, and Using Deep Learning for Early Sepsis Prediction," 2019 Computing in Cardiology (CinC), 2019, pp. Page 1-Page 4, doi: 10.23919/CinC49843.2019.9005940.

[13] Reyna, M., Josef, C., Jeter, R., Shashikumar, S., Moody, B., Westover, M. B., Sharma, A., Nemati, S., & Clifford, G. (2019). Early Prediction of Sepsis from Clinical Data -- the PhysioNet Computing in Cardiology Challenge 2019 (version 1.0.0). PhysioNet.

[14] M. Shahul and K. P. Pushpalatha, "Machine Learning Based Analysis of Sepsis: Review," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.399.

[15] M. J. Pettinati, G. Chen, K. S. Rajput and N. Selvaraj, "Practical Machine Learning-Based Sepsis Prediction," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 4986-4991, doi: 10.1109/EMBC44109.2020.9176323.

[16] A. Shankar, M. Diwan, S. Singh, H. Nahrpurawala and T. Bhowmick, "Early Prediction of Sepsis using Machine Learning," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 837-842, doi: 10.1109/Confluence51648.2021.9377090.

[17] F. Mahmud, N. S. Pathan and M. Quamruzzaman, "Early detection of Sepsis in critical patients using Random Forest Classifier," 2020 IEEE Region 10 Symposium (TENSYMP), 2020, pp. 130-133, doi: 10.1109/TENSYMP50017.2020.9231011