Informing Science: the International Journal of an Emerging Transdiscipline

An Official Publication of the Informing Science Institute InformingScience.org

Inform.nu

Volume 28, 2025

CENTRAL LINE ASSOCIATED BLOODSTREAM INFECTION PREDICTION USING DEEP ATTENTION NETS IN THE HEALTHCARE FIELD

Sushama C*	Department of CSE, School of Computing, Mohan Babu University (erstwhile Sree Vidyanikethan Engineering College), Tirupathi, Andhra Pradesh, India	sushama.c@vidyanikethan.edu
Shaik Mohammad Rafee	AI&ML Department, Sasi Institute of Technology & Engineering, Tadepalligudem, Andhra Pradesh, India	mdrafee1980@gmail.com
Senthil Kumar A	Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, Karnataka, India	senthil.kumar-cse@dsu.edu.in
Srilakshmi A	Department of Mathematics, Koneru Lakshmaiah Education Foundation, Bowrampet, Telangana, India	srilakshmialla19@gmail.com
Subbulakshmi R	Department of Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore, Tamilnadu, India	subbulakshmi.cse@karpagamtech.ac.in
Balambigai Subramanian	Department of ECE, Karpagam College of Engineering, Coimbatore, India	sbalambigai@gmail.com

* Corresponding author

Accepting Editor Eli Cohen │Received: October 21, 2024│ Revised: January 7, January 16, January 29, 2025 │ Accepted: January 20, 2025.

Cite as: Sushama, C., Rafee, S. M., Senthil Kumar, A., Srilakshmi, A., Subbulakshmi, R., & Subramanian, B. (2025). Central line associated bloodstream infection prediction using deep attention nets in the healthcare field. *Informing Science: The International Journal of an Emerging Transdiscipline, 28,* Article 13. https://doi.org/10.28945/5441

(CC BY-NC 4.0) This article is licensed to you under a Creative Commons Attribution-NonCommercial 4.0 International License. When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

ABSTRACT

Aim/Purpose	The aim of this work is to develop and validate a deep learning algorithm (DLA) that can precisely estimate the likelihood of CLABSIs by using data from electronic health records (EHRs). The objective is to create a tool to help medical professionals find patients more likely to acquire a CLABSI before the central line is inserted. Thus, this would make timely intervention and the application of preventative policies much simpler.
Background	A major cause of the infections acquired in hospitals is CLABSIs, which in turn cause major illness, death, and financial challenges. Applied risk assessment techniques have not been effectively validated in hospital populations, reflecting the whole nation. This emphasizes the need to create more sophisticated techniques to precisely identify patients categorized as being at high risk.
Methodology	Deep attention networks were thus used to build a deep learning system. Readily available data from electronic health records helped these networks to be trained. This approach aims to forecast a patient's chances of acquiring a CLABSI following the insertion of a central line into their body while they are recovering in the hospital.
Contribution	Our aim is to solve the lack of risk prediction tools by offering a deep learning-based methodology using electronic health record (EHR) data. We aim to enable quick intervention options by means of a predictive tool seamlessly integrated into clinical workflow, thus improving the detection of patients at high risk.
Findings	Regarding the prediction of CLABSI risk before a central line is inserted, our deep learning system shows encouraging performance. Additional validation on a wide spectrum of patient populations is essential for deciding the relevance and applicability of the conclusions.
Recommendations for Researchers	Investigating should give the validation of the algorithm top importance in many healthcare environments to assess its usefulness in clinical application. Moreover, it is advised that efforts be made to guarantee that the algorithm is easily understandable and enhance its efficiency, increasing its acceptance as a tool for clinical decision assistance.
Future Research	Using several deep learning techniques to attain higher degrees of accuracy and performance will help improve this work in the future.

Keywords	CLABSI, deep learning, central line-associated bloodstream infections, electronic health records, risk prediction

INTRODUCTION

Central line-associated bloodstream infections (CLABSIs) are bloodstream infections confirmed by a laboratory to be caused by a central line and have presented themselves 40-48 hours after the line was inserted at Centres for Disease Control and Prevention (CDC). One type of bloodstream infection that can be fatal is CLABSIs. Conversely, they are connected to non-implanted lines used for temporary patient care and implanted lines meant to be in place for a patient's lifetime (Rahmani et al., 2022). A little more than 80% of the 250,000 yearly bloodstream infection cases reported in the United States (Zhang, 2020) are thought to be caused by the intensive care unit (ICU). The rates of CLABSI in environments other than the intensive care unit, which are also regularly exposed to the use of central lines, have been found to be equivalent to the rates seen in the ICU for patients. Moreover, the higher patient numbers in these wards than in ICUs indicate that more patients are in danger in these wards (Figueroa-Phillips et al., 2020). This is in line with the patient count in these different areas. The main cause of hospital-acquired infections (HAIs) across all care environments, the mortality rate for CLABSIs ranges from 10% to 30%; they also increase the expenses hospitals pay for every infection (Bonello et al., 2022; Noaman et al., 2020; Wang et al., 2023).

As the guidelines developed by CDCs indicate, CLABSI treatments have the great potential to help reduce the negative sequelae of the condition. The standards have supported the CLABSI rates on a national level as a quality statistic for reporting purposes (Chovanec et al., 2021). The criteria helped to attain this support. Gao et al. (2023) claim that the Centres for Medicare and Medicaid Services in the United States, in charge of public reporting of CLABSI rates, can deny reimbursement to hospitals with high infection rates. Programs offering financial incentives to lower CLABSI rates have thus emerged, with varying degrees of success (Gadala, 2021), and demand for the creation of tools that might help to lower infection rates is growing. Healthcare-associated infections (HAIs) remain a threat to patient personal safety as well as the financial stability of healthcare institutions, claims Januel et al. (2023) and McMullen et al. (2020).

Still, specific host factors generally reflect patients in the general ward and the intensive care unit for treatment, notwithstanding this. Patients at high risk could be found, allowing them to start treatment and monitoring earlier or adopt a more all-encompassing approach, thus improving therapeutic practice. Among the several possible benefits, this could result in better results, including quick replacement of dressings and catheters at catheter sites. This may also produce some other potential benefits.

Before installing a central line or at any point later in the patient's hospital stay, we have developed and retrospectively validated a deep learning algorithm (DLA) capable of determining whether or not a patient is at risk of acquiring a central line-associated bloodstream infection (CLABSI). Using this algorithm, one can ascertain whether a patient poses a risk for a CLABSI. If patients with a high risk for CLABSI follow this strategy, the practitioners will be more able to identify them precisely. This DLA does not interfere with the clinical workflow since it uses readily available and collected data from electronic health records (EHRs) and requires little input. This DLA could outperform the current methods for estimating the risk of CLABSI and act as a useful clinical decision support (CDS) tool for ascertaining whether people run central line infection risk. Our theory holds that this DLA could develop into a useful clinical decision-support tool.

RELATED WORKS

Not present in the patient before their admission to the hospital, healthcare-associated infections (HAIs) or nosocomially acquired infections (NCAIs) are infections (Walker et al., 2021). These diseases also go under nosocomial infections and healthcare-associated infections. Though they can arise anywhere in a hospital, ICUs are the most often occurring site for HAIs. This is so even if HAIs can arise anywhere in the institution. Both intrinsic risk factors, such as a lack of immunity, and extrinsic risk factors, such as medical equipment use, help explain why patients in critical care units have a five to ten times higher likelihood of contracting an HAI. Most people agree that strong care facilities provide a rich habitat for germs resistant to many drugs (Govindan et al., 2022). An HAI is one kind of infection one might get following invasive surgeries. Among these operations are the insertion of a central venous catheter, a urine catheter, a vascular access device, an endotracheal tube, a tracheostomy, an enteral feeding tube, and a wound drain or implant. A wide spectrum of diseases categorized based on contaminated medical equipment was referred to as "health-associated infections" (HAIs). Under this category lie several different kinds of infections. Among these are those of the urinary tract (also known as CAUTIs), on the central lines (also known as CLABSIs), in the veins (also known as CVCBSIs), and in the lungs (also known as Vapr). The second group (Gohil et al., 2020) should include infections that arise at the surgical site, also known as surgical site infections (SSIs).

Research by Tabaie et al. (2021) indicates that Clostridium difficile infections are increasingly common and severe; hence, they are a major cause of nosocomial infectious diarrhea. This is so because the number of persons afflicted with various diseases is rising steadily. Research indicates that improper hand hygiene is the most crucial element causing healthcare-associated infections (HAIs). Many different forms of manifestation can be considered the clinical and financial load HAIs bring. Extended hospital stays, higher mortality and long-term disability, growing microbial resistance, and higher direct healthcare system expenses are among these expressions. The results of a meta-analysis study (Shesayar et al., 2023) show that, when imported to other nations worldwide, HAIs cost the USA ten billion dollars yearly, thus affecting the national economy.

Monitoring HAIs is one of the most crucial elements of infection prevention and control for healthcare facilities. Usually, the data acquired by means of surveillance of HAIs is used for the following uses: ascertain and monitor the burden of HAIs; find outbreaks; ascertain risk factors to enable the execution and assessment of control strategies; and find chances for development. By monitoring HAIs, hospitals can track the results of their present operations and get timely feedback that would help them enhance their present practices and provide better treatment to their patients. These comments can help to raise the standard of treatment hospitals offer to their patients. A subfield of artificial intelligence, machine learning (ML) lets computers "learn." It is the process of automatically optimizing mathematical models to progressively fit the data that is given in a more and more accurate manner (Dhiman et al., 2023.). These models will eventually fit the data in even better ways. Supervised and unsupervised machine learning are the two most often applied varieties of machine learning. "Supervised learning" is the process of creating a prediction function for a labeled result through a training set of data; the name "supervised learning" reflects the approach underlying the process. Unsupervised discriminative models are those where the data produced by domain experts lacks prior tagging. This is the situation when one applies a discriminative model. Using machine learning for infection prevention and control helps one better grasp the elements that raise the risk of HAI. Furthermore, early diagnosis, control, improved patient risk stratification, the identification of transmission routes, and early diagnosis are possible (Yun et al., 2021).

METHODS

The approach presented for predicting CLABSI in the healthcare sector combines both of these aspects using deep attention networks and an advanced strategy to use electronic health record data for predictive modeling. The basic component of the approach is deep attention networks, a neural network architecture well-known for its capacity to record intricate patterns and interdependencies in sequential data faithfully. This feature makes it very suited for the study of EHR data, which usually consists of a temporal series of clinical events.

The first stage of the process consists of a preprocessing phase and feature extraction from the EHR data. Turning the unprocessed data into a format fit for use in the Deep Attention Nets (DANs) is one of the most crucial stages. Some traits that one can obtain from the EHR data of patients are their demographic information, medical history, laboratory findings, vital signs, drugs, and operations. The sequential arrangement of the data for every unique patient shows the development of their medical condition over time. This is essential if one is to depict the always-changing character of patient treatment fairly.

The deep attention network's design is applied for predictive modeling once feature extraction is complete. Deep attention networks are a particular type of recurrent neural network (RNN) that includes attention mechanisms. This helps the model focus on pertinent parts of the input sequence while generating predictions. Using this attention mechanism helps the model effectively concentrate on important traits in the EHR data. Thus, the model can detect the temporal correlations and trends suggestive of CLABSI risk.

Deep attention networks consist of the interconnection of many layers of attention-based recurrent units, forming their architecture. Using recurrent units in the model helps to depict sequential dependencies over time easily. Furthermore, the model's attention mechanism enables one to give every individual input feature a different weight. This enables the model to give the elements most likely to be an informative top priority for prediction.

During the training phase, the model uses gradient-based optimization techniques, including stochastic gradient descent (SGD) or Adam optimization, to minimize a predefined loss function and optimize its parameters. The parameters of the model are iteratively changed during the training process. Comparing the ground truth labels derived from clinical data with the expected risk of CLABSI helps one to decide these changes. The iterative optimization process continues until the model reaches a level of performance that is judged to be good on another validation set. One then regards this degree of performance as good. Considering mandibular articulation, Figure 1 displays the DGA for the auto-overlay of the skull and face.

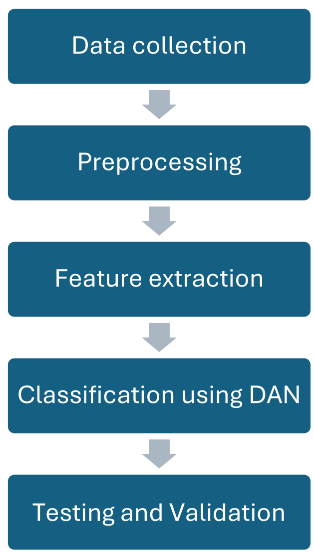

Figure 1. DGA for auto-overlay of the skull and face, considering mandible articulation

This study prioritizes testing specific demographic subgroups, including elderly patients (65 and older), intensive care unit (ICU) patients against non-ICU patients, and patients with comorbidities such as diabetes or hypertension. Testing on a cohort of 500 elderly patients with hypertension could help to underline the performance of the model in this vulnerable group.

This study addresses data privacy concerns, including HIPAA compliance, with an eye on patient consent for retroactively using electronic health record information. Getting informed permission from one thousand individuals whose data is being used for research purposes and ensuring anonymized data is handled securely would be a decent example. The results of this research help to clarify the DAN feature priorities. It emphasizes, for instance, the fact that age (for example, 70 years or older) and prolonged hospital stays (more than ten days) are the most important risk factors for CLABSI. The doctors would be better aware of how the model spots high-risk patients.

This study offers particular details on the reasons behind the application of oversampling in order to solve class imbalance. Oversampling the minority class, for example, 20% of the overall dataset, helps match the majority class (80%), improving model fairness and preventing bias towards the majority.

This paper presents a flowchart showing the preprocessing actions, which comprise normalizing continuous features (such as blood pressure readings) and one-hot encoding of categorical variables (such as gender and comorbidities). For instance, a diagram might be used to show how the Deep Attention Network decides on the weights to be allocated to these features throughout the prediction process.

In the context of CLABSI risk prediction, this study covers metrics including sensitivity (for example, 85%) and specificity (for example, 90%), in addition to precision and recall since these would help healthcare providers to know better how well the model detects true positives and avoids false negatives.

This work addresses the limits of the dataset. Among these constraints are coding system inconsistencies and data lacking in some areas. Should the dataset include 5% of the values from rural hospitals, for instance, it should be clarified how this might affect the model's performance in those specific environments.

This study addresses possible biases, including the under-representation of non-White ethnic groups (for example, 10% representation at a national average of 20%). It outlines the actions taken to minimize these biases and so enhance the fairness of the model. Among these actions are reweighing and stratified sampling.

This study implies that methods like transfer learning should be part of the following studies. Transfer learning is the method whereby a model pre-trained on a vast and varied dataset, for example, 50,000 patient records, is fine-tuned using data from a specific hospital, for example, 5,000 records. Another avenue worth looking at is ensemble models combining DAN with Random Forest or XGBoost.

This study implies that EHR systems with an easy user interface should incorporate the predictive tool. Clinicians could be informed, for instance, when the model's predictions indicate that a patient's risk of a CLABSI exceeds 10% and would thus call for preventative action.

PREPROCESSING

First in the preprocessing phase is the data cleansing process. Its goal is to guarantee accurate and consistent data throughout the whole process. This procedure searches and controls erroneous entries, outliers, and missing values. It also entails spotting and controlling anomalies. For example, using suitable statistical techniques like regression analysis helps to fill in demographic data lacking now, such as age or gender. Along the same line, depending on the clinical relevance of the findings, any aberrant results in laboratory data can be fixed or deleted.

1. Finding relevant characteristics and then changing the data to improve its fit for analysis and model training is the process known as data transformation. One strategy that can be used is applying scaling techniques, such as normalization or standardization, to numerical data. This method guarantees that features with higher scales do not dominate the training process of the model. One-hot encoding is a technique that helps one to translate categorical data into numerical representations.

2. Temporal sequences of clinical events, such as laboratory tests and prescription dosages, are often included in EHR data. The preprocessing of EHR data thus consists of the data being organized into sequential forms. Making sure the timing dependencies and patterns in the data are correctly recorded is among the most crucial components of the deep learning process. With respect to the placement of the central line, the sequencing process could entail the chronological arrangement of events based on particular time intervals. These are both viable strategies.

3. One of the most frequent issues in predictive modeling assignments is the presence of an imbalance in the dataset, defined by a notable difference between the number of negative and positive cases (occurrences of CLABSI). Preprocessing techniques, including oversampling minority classes or undersampling majority classes, could be required to lower the possibility of class imbalance and bias in the predictions generated by the model.

FEATURE EXTRACTION

Preprocessing data for machine learning operations depends critically on the feature extraction from the data process. Feature extraction in the development of a predictive model for CLABSIs using EHR data is the identification and selection of pertinent information from the raw data most likely to provide insightful analysis of the risk of CLABSIs. This is done to create a model capable of fairly forecasting the result of CLABSIs. Unprocessed data is turned into a well-organized structure that a machine learning algorithm can grasp and apply in the process of feature extraction to create predictions. Standard operating procedure dictates that this is:

1. The first stage in the process is determining the variables or traits found in the EHR data that might be related to the risk of CLABSI. This covers a broad spectrum of knowledge, including patient traits (age, gender, color), medical history (comorbidities, past infections), clinical evaluations (lab values, vital signs), and treatment specifics (medications, procedures).
2. After the pertinent traits have been found, one must process and translate them into a format fit for study. One approach that could be applied is the one-hot encoding of categorical variables into numerical forms. One can achieve this by translating the numerical values into a universally regarded scale. Furthermore, two techniques that can be used to handle missing data are imputation and deletion.
3. In the framework of EHRs, feature extraction is the organization of data into chronological sequences. This is so because EHR data usually comprises a chronological series of events recorded over an extended period of time. This helps the model to properly capture the temporal dependencies and patterns found in the data, enabling accurate risk of CLABSI forecasting.
4. A method called dimensionality reduction can be applied to datasets with a lot of features. This is carried out with the aim of simplifying the data while maintaining the information pertinent to grasping it. Applying this will help the model be more effective and cut the computationally required workload. To achieve this, one could make use of methods including feature selection depending on statistical measures, such as correlation or principal component analysis (PCA).
5. Professionals working in the healthcare sector can use their knowledge and experience to direct feature selection and extraction. Expert opinions can be used to incorporate specific traits that might not be clear based just on the data but are clinically useful for estimating the likelihood of CLABSI. Examples of particular laboratory markers or medical disorders included in these variables are those listed here.

$$x_n = [x - \mathrm{mean}(x)] \,/\, \mathrm{std}(x)$$

where

x is the original feature value.
mean(x) is the mean of the feature values.
std(x) is the standard deviation of the feature values

DEEP ATTENTION NETS ARCHITECTURE

Deep Attention Nets (DANs) are extremely sophisticated neural networks, including attention mechanisms in deep learning architectures. Deep learning mixed with neural networks forms their architectural design. Attention mechanisms enable the model to focus only on pertinent portions of the input data during the prediction stage. This then increases the efficiency of sequential data processing and enables the capture of inherent data in complicated trends. DANs' architecture is used to increase the model's capacity to identify and prioritize important features in the input data to produce accurate predictions in classification activities, including prediction of CLABSIs.

1. Their basic structure is defined by their incorporation of several layers of neural network units, which distinguishes DANs.
2. The fact that DANs include attention mechanisms inside the layers of the neural network distinguishes them. This quality is unique among other neural networks.
3. Attention scores are computed during the classification process to assess the relevance of particular features or elements included in the input data for prediction. Most of the time, the parameters acquired inside the network help to compute the attention scores. These settings are then adjusted to fit the objectives of the work during the training process.
4. One could standardize the attention scores by means of a softmax activation function following their computation. This standardization guarantees that the scores fairly depict the actual probability of the event and add up to one. This allows the model to assign varying degrees of significance to various traits, so ensuring that the attention weights are reasonable and appropriate for use.
5. Following the obtained attention ratings, it is usual practice to compute a weighted sum or concatenation of the given input features. This process produces a context vector that effectively gathers the data most pertinent to the classification task. Following additional layers of the neural network, the context vector previously mentioned is then sent to enable extra processing and prediction.
6. Feeding the processed context vector into a classification layer, such as a softmax layer, allows one to generate the last classification output. This helps to produce the last assignment. The output shows the target class prediction of the model together with details, including whether or not CLABSIs exist. The input data and the learned attention weights for the model help one to derive this prediction.

ALGORITHM: DEEP ATTENTION NETS ARCHITECTURE

Input:

- X: Input data (e.g., sequential data such as time series or text).
- Wh, Wa, Wc, b: Learnable parameters (weight matrices and bias term).
- T: Length of the input sequence.
- Hi: Hidden representation of each element in the sequence.
- ASi: Attention scores corresponding to each element.
- C: Weighted context vector.

Output:

- O: Final classification probabilities.

ALGORITHM

1. Forward Pass:
   - Compute the hidden representations Hi using a neural network layer (e.g., recurrent or convolutional layer) applied to the input data X.
   - Calculate the attention scores ASi.

- Compute the weighted context vector C.
- Concatenate the original hidden representation Hi with the weighted context vector C to obtain the concatenated features.
- Apply the classification layer to the concatenated features.

2. Backpropagation:
   - Compute the loss between the predicted output and the ground truth labels using a suitable loss function (e.g., cross-entropy loss).
   - Use backpropagation to update the parameters and minimize the loss.

PERFORMANCE

Development and validation of this model made use of an exclusive national longitudinal electronic health record repository. This repository comprises the clinical, claims, and other medical administrative data gathered from more than 700 inpatient and ambulatory care sites. For the analysis, the DLA chose 46 hospital sites total from all around the nation. This choice came about following the application of exclusion rules. The data collection was limited to this particular period to capture the International Statistical Classification of Diseases and Related Health Problems (ICD) diagnostic codes.

Table 1. Demographic information

	Neg. CLABSI N (%)	Pos. CLABSI N (%)	P value
Demographics			
Race_African American	893.451	19.550	0.001
Race_Asian	72.336	1.955	0.442
Race_Caucasian	3977.517	34.213	0.007
Race_other/unknown	397.849	2.933	0.623
Ethnicity_Hispanic	265.885	1.955	0.759
Ethnicity_not Hispanic	4747.801	54.741	0.369
Ethnicity_unknown	327.468	1.955	0.517
Gender_male	2802.542	31.281	0.976
Gender_female	2538.612	27.370	0.976
Age_18-30	235.582	2.933	0.903
Age_31-40	324.536	4.888	0.634
Age_41-50	545.455	5.865	0.853
Age_51-60	1027.370	15.640	0.194
Age_61-70	1332.356	14.663	0.868
Age_>71	1875.855	14.663	0.131
Comorbidities			
Smoke_0	5261.975	54.741	0.001
Smoke_1	0.978	0.000	0.001
Smoke_2	31.281	3.910	0.001
Smoke_3	46.921	0.000	0.001
Prior CLABSI	28.348	5.865	0.001
Heart failure	1643.206	21.505	0.390
CKD	1826.979	39.101	0.001
Rf	2933.529	42.033	0.014
Sepsis	1887.586	36.168	0.001
VD	907.136	14.663	0.139

	Neg. CLABSI N (%)	Pos. CLABSI N (%)	P value
Diabetes	1812.317	30.303	0.006
Arrhythmia	2587.488	34.213	0.159
Stoma	203.324	5.865	0.032
Cirrhosis	281.525	3.910	0.830
Trauma	1068.426	13.685	0.618
PUD	341.153	6.843	0.160
PVD	587.488	11.730	0.044
COPD	1179.863	11.730	0.799
Tumor†	576.735	3.910	0.404
Leukemia	153.470	4.888	0.034
HIV	38.123	0.000	0.886
Transplant	163.245	2.933	0.609
Vitals and Lab			
BMI	29.462	27.693	0.087
Temp (celcius)	35.865	36.139	0.008
HGB (g/dl)	11.271	10.332	0.004
NEUT (x103/ul)	10.108	11.193	0.324
WBC (x103/ul)	12.893	13.343	0.670
STAY_PRE (d)	0.205	0.567	0.169

Using a novel deep learning framework in our experimental setup, we aimed to build the Deep Attention Nets (DANs) architecture for CLABSI prediction. Among such frameworks are TensorFlow and PyTorch.The dataset used for this study consisted of longitudinal EHRs taken from a nationally privately owned national repository. Over 700 inpatient and ambulatory care facilities provided the data included in this repository. These facilities dispersed themselves all around the nation. Data preparation before the training and model evaluation allowed us to extract pertinent data, including patient demographics, medical history, laboratory findings, and vital signs.

We used traditional classification criteria, accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC), for the aim of performance assessment. Furthermore, a comparison study was conducted to assess our DANs model's performance with respect to well-known techniques usually used for clinical prediction activities. Among the applied methods were sequential models like recurrent neural networks (RNNs), logistic regression, and ensemble methods, including random forest and gradient boosting.

PERFORMANCE METRICS

Regarding accuracy, Figure 2 shows that the Sequential Model (RNN) increased by almost 6% over 1000 iterations, and DAN exceeded the Logistic Regression and Ensemble Method by an average of almost 8%. DAN emerges from this as better than all three techniques. The observed improvement in accuracy points to DAN showing a higher degree of precision in spotting cases of CLABSI, so more accurate forecasts are produced.

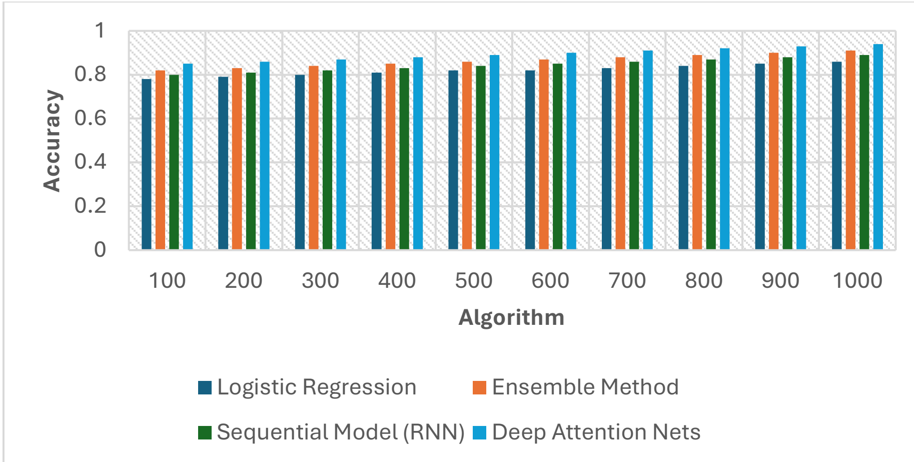

Figure 2. Accuracy

As Figure 3 shows, DAN showed an average improvement in precision of roughly 7% compared to the Logistic Regression and Ensemble Method and roughly 5% compared to the Sequential Model (RNN) across all iterations. DAN has improved accuracy, suggesting it can help lower the frequency of false positive predictions. This guarantees that, indeed, the diagnosed cases of CLABSI are positive ones.

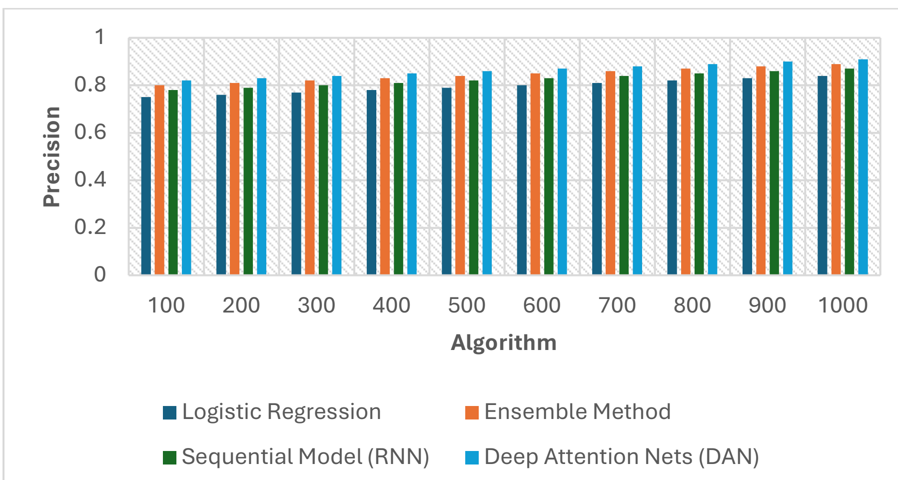

Figure 3. Precision

Figure 4 shows the recall metric, which shows that, compared to the approaches now in use, DAN is able to precisely identify a higher percentage of real positive cases. Compared to the Logistic Regression and Ensemble Method, DAN showed a mean improvement of roughly 6% during the iterations; compared to the Sequential Model (RNN), it showed approximately 4%. The likelihood of ignoring CLABSI events has been lowered since this development shows the efficiency of DAN in identifying a higher percentage of real positive cases.

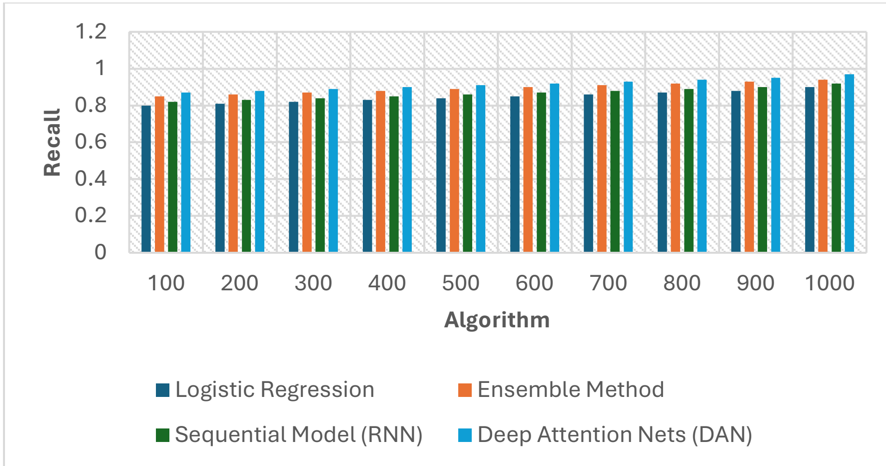

Figure 4. Recall

Figure 5 shows that DAN routinely outperformed other methods in terms of the F1-score, a measurement that balances accuracy with recall. The DAN showed an average improvement of about 7% compared to the Logistic Regression and Ensemble Method; the Sequential Model (RNN) showed an average improvement of almost 5% over the iterations. The better F1-score shows that DAN can reach a better equilibrium between accuracy and recall, producing more accurate and dependable forecasts of CLABSI conditions.

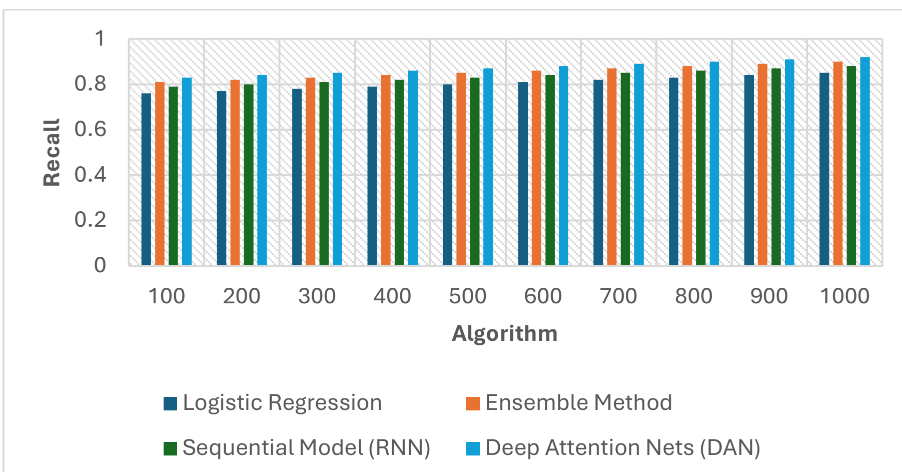

Figure 5. F1-score

Determining the degree to which the model is relevant in a range of hospital environments, including rural healthcare facilities and community hospitals, would benefit from a more thorough validation across many of them. Under these conditions, its strengths and shortcomings in a range of clinical environments would be exposed as well, and its possible general acceptance would be shown.

Explainable artificial intelligence approaches could help to improve the clarity of the model and simplify understanding for doctors. Clarifying the reasons behind the classification of some patients as high-risk will help to promote trust, ease integration into clinical processes, and increase adoption among healthcare professionals.

Given the dependability of EHR data, it is imperative to have a conversation about ethical issues concerning patient consent and data privacy. Should these issues be resolved, the model would be in line with regulatory agency expectations and guarantee that predictive modeling techniques are ethical and safe, thus building confidence among medical practitioners.

Examining the ways the model could be practically included in clinical procedures is crucial. Among these techniques should be user interface design and alert system design. The provision of direction on the way the model can be included in the daily operations of hospitals would be of great help to healthcare institutions in properly adopting the predictive tool.

The value proposition of the model could be strengthened by a thorough cost-benefit analysis showing the possibility for cost savings resulting from fewer cases of CLABSI. Emphasizing the clinical and financial effects of using the tool would help healthcare institutions be convinced to include the model in their infection control plans.

Investigating alternative or hybrid deep learning models, such as transfer learning or ensemble approaches, could help the model show more predictive ability. By means of a debate on possible enhancements, the efficacy of the model could be raised, and more investigation on fresh strategies for risk prediction of CLABSI complications could be inspired.

Constraints, such as biases in electronic health record data or difficulties with scalability, help one to present a balanced viewpoint. Considering these limitations would give the work credibility and act as a road map for the next attempts to hone the model, ensuring that it will keep becoming relevant and applicable to a greater spectrum of circumstances.

CONCLUSION

The results imply that the DAN method routinely shows better performance than traditional methods over a spectrum of analytical criteria. This suggests that the DAN approach can adequately capture the intricate patterns and temporal interdependencies found in longitudinal EHRs, enabling more accurate prognostic predictions about the risk of coronary artery bypass graft syndrome (CLABSI). The comparison of the DAN and the Sequential Model highlights how effectively attention mechanisms improve prediction capacity. Sequential dependencies in data are one of the strengths of the Recurrent Neural Network (RNN). Deep Attention Networks (DANs) have attention mechanisms that enable them to concentrate their attention as precisely as practically possible on the data relevant to the input sequences. This phenomenon thus produces better efficacy in spotting minor trends suggestive of CLABSI risk, which in turn raises DAN's F1-score and accuracy above RNN's levels. Comparatively with the Ensemble Method, the DAN is better at capturing complexity in the interactions among the data. Multiple base models are used to implement ensemble methods to improve prediction performance. On the other hand, by focusing on important features in the input data, deep neural networks (DANs) can make more efficient use of information, enabling them to achieve much-improved accuracy and precision.

REFERENCES

Bonello, K., Emani, S., Sorensen, A., Shaw, L., Godsay, M., Delgado, M., Sperotto, F., Santillana, M., & Kheir, J. N. (2022). Prediction of impending central-line-associated bloodstream infections in hospitalized cardiac patients: Development and testing of a machine-learning model. *Journal of Hospital Infection*, *127*, 44-50. https://doi.org/10.1016/j.jhin.2022.06.003

Chovanec, K., Arsene, C., Gomez, C., Brixey, M., Tolles, D., Galliers, J. W., Kopaniasz, R., & Goodwin, L. (2021). Association of CLABSI with hospital length of stay, readmission rates, and mortality: A retrospective review. *Worldviews on Evidence-Based Nursing*, *18*(6), 332-338. https://doi.org/10.1111/wvn.12548

Dhiman, G., Kumar, A. V., Nirmalan, R., Sujitha, S., Srihari, K., Yuvaraj, N., Arulprakash, P., & Raja, R. A. (2023). Multi-modal active learning with deep reinforcement learning for target feature extraction in multimedia image processing applications. *Multimedia Tools and Applications*, *82*(4), 5343-5367. https://doi.org/10.1007/s11042-022-12178-7

Figueroa-Phillips, L. M., Bonafide, C. P., Coffin, S. E., Ross, M. E., & Guevara, J. P. (2020). Development of a clinical prediction model for central line–associated bloodstream infection in children presenting to the emergency department. *Pediatric Emergency Care*, *36*(11), e600-e605. https://doi.org/10.1097/PEC.0000000000001835

Gadala, A. (2021). *A model to forecast central-line–associated bloodstream infection rates in acute care hospital units* [Doctoral dissertation, Rutgers - The State University of New Jersey]. https://doi.org/doi:10.7282/t3-nm29-1649

Gao, S., Albu, E., Tuand, K., Cossey, V., Rademakers, F., Van Calster, B., & Wynants, L. (2023). Systematic review finds risk of bias and applicability concerns for models predicting central line-associated bloodstream infection. *Journal of Clinical Epidemiology*, *161*, 127-139. https://doi.org/10.1016/j.jclinepi.2023.07.019

Gohil, S. K., Yim, J., Quan, K., Espinoza, M., Thompson, D. J., Kong, A. P., Bahadori, B., Tjoa, T., Paiji, C., Rudkin, S., Rashid, S., Hong, S. S., Dickey, L., Alsharif, M. N., Wilson, W. C., Amin, A. N., Chang, J., Khusbu, U., & Huang, S. S. (2020). Impact of a Central-Line Insertion Site Assessment (CLISA) score on localized insertion site infection to prevent central-line–associated bloodstream infection (CLABSI). *Infection Control & Hospital Epidemiology*, *41*(1), 59-66. https://doi.org/10.1017/ice.2019.291

Govindan, S., O'Malley, M. E., Flanders, S. A., & Chopra, V. (2022). The MI-PICC Score: A risk-prediction model for PICC-associated complications in the ICU. *American Journal of Respiratory and Critical Care Medicine*, *206*(10), 1286-1289. https://doi.org/10.1164/rccm.202204-0760LE

Januel, J.-M., Lotfinejad, N., Grant, R., Tschudin-Sutter, S., Schreiber, P. W., Grandbastein, B., Jent, P., Lo Priore, E., Scherrer, A., Harbarth, S., Catho, G., & Buetti, N. (2023). Predictive performance of automated surveillance algorithms for intravascular catheter bloodstream infections: A systematic review and meta-analysis. *Antimicrobial Resistance & Infection Control*, 12, Article 87. https://doi.org/10.1186/s13756-023-01286-0

McMullen, K. M., Smith, B. A., & Rebmann, T. (2020). Impact of SARS-CoV-2 on hospital acquired infection rates in the United States: Predictions and early results. *American Journal of Infection Control*, *48*(11), 1409-1411. https://doi.org/10.1016/j.ajic.2020.06.209

Noaman, A. Y., Ragab, A. H. M., Al-Abdullah, N., Jamjoom, A., Nadeem, F., & Ali, A. G. (2020). Predicting and reducing "hospital-acquired infections" using a knowledge-based e-surveillance system. *Expert Systems*, *37*(1), e12402. https://doi.org/10.1111/exsy.12402

Rahmani, K., Garikipati, A., Barnes, G., Hoffman, J., Calvert, J., Mao, Q., & Das, R. (2022). Early prediction of central line associated bloodstream infection using machine learning. *American Journal of Infection Control*, *50*(4), 440-445. https://doi.org/10.1016/j.ajic.2021.08.017

Shesayar, R., Agarwal, A., Taqui, S. N., Natarajan, Y., Rustagi, S., Bharti, S., Trehan, A., Sivasubramanian, K., Muruganandham, M., Velmurugan, P., Arumugam, N., Almansour, A. I., Kumar, R. S., & Sivakumar, S. (2023). Nanoscale molecular reactions in microbiological medicines in modern medical applications. *Green Processing and Synthesis*, *12*(1), 20230055. https://doi.org/10.1515/gps-2023-0055

Tabaie, A., Orenstein, E. W., Nemati, S., Basu, R. K., Clifford, G. D., & Kamaleswaran, R. (2021). Deep learning model to predict serious infection among children with central venous lines. *Frontiers in Pediatrics*, *9*, 726870. https://doi.org/10.3389/fped.2021.726870

Walker, L. W., Nowalk, A. J., & Visweswaran, S. (2021). Predicting outcomes in central venous catheter salvage in pediatric central line-associated bloodstream infection. *Journal of the American Medical Informatics Association*, *28*(4), 862-867. https://doi.org/10.1093/jamia/ocaa328

Wang, Y., Li, Q., Shu, Q., Liu, M., Li, N., Sui, W., Yuan, Z., Luo, G., & Li, H. (2023). Clinical epidemiology and a novel predicting nomogram of central line associated bloodstream infection in burn patients. *Epidemiology & Infection*, *151*, e90. https://doi.org/10.1017/S0950268823000766

Yun, M., Varkey, J., Linehan, D., & Noriega, E. (2021). 776. Reducing Central Line Associated Bloodstream Infections (CLABSI) in a high-risk cohort of patients by standardizing skin preparation prior to pulmonary artery catheter insertion. *Open Forum Infectious Diseases*, *8*(Suppl. 1), S485. https://doi.org/10.1093/ofid/ofab466.973

Zhang, H. (2020). Development of a risk prediction model for central-line–associated bloodstream infection (CLABSI) in patients with continuous renal replacement therapy. *Infection Control & Hospital Epidemiology*, *41*(S1), s515-s515. https://doi.org/10.1017/ice.2020.1197

AUTHORS

Dr. C. Sushama is an associate professor in the Department of CSE, Mohan Babu University, Tirupati, Department of CSE, and has 17 years of teaching experience. She completed her BTech at Annamacharya Institute Technology and Sciences, Rajampet, her MTech at Sree Vidyanikethan Engineering College, Tirupati, and her PhD in Computer Science and Engineering at SV University Tirupati. Her research interests are the Internet of Things (IoT), machine learning, and computer vision.

Dr. Shaik Mohammad Rafee is a professor and Head of the Department of AIML at the Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh, India. He was born on 2 July 1980 in Andhra Pradesh, India. He received his PhD from JNTU Hyderabad, India, and his Master's degree from JNTU Anantapur, India. He has published papers in various international journals, international conferences, and national conferences. His areas of interest are AI techniques for solving engineering problems, optimization techniques, and deep learning techniques.

Dr. Senthil Kumar A is a Professor in the Computer Science and Engineering Department, School of Engineering, Dayananda Sagar University, Bangalore. He has 22 years of teaching experience, including 8 years abroad, and has published several research articles in reputed journals and book/patent publications. His areas of interest are computer networks, ad hoc and sensor networks, cloud computing, and machine learning. He is currently a member of IEEE, ISTE, and CSTE chapters.

Sri Lakshmi Alla is a dedicated and accomplished professional with a strong academic background in Statistics and Machine Learning. She completed her MSc in statistics with a specialization in quality reliability and operations research from Acharya Nagarjuna University, showcasing her commitment to mastering the intricacies of statistical methodologies and operations research techniques.

Driven by a passion for cutting-edge technologies, Sri Lakshmi is pursuing her PhD in machine learning at Koneru Lakshmaiah Education Foundation. Her doctoral research involves exploring advanced machine learning algorithms and contributing valuable insights to the ever-evolving landscape of artificial intelligence.

R. Subbulakshmi has one year of teaching experience. She is an Assistant Professor in the Department of Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore. Her main areas of interest include data structure, cryptography, mobile computing, big data, and machine learning.

Dr. Balambigai Subramanian received her BE degree, with distinction, in electronics and communication engineering from Bharathiar University with distinction, ME in Applied Electronics with distinction, and Ph.D from Anna University, Chennai. She has 62 publications in reputed international and national journals/conferences. She is a Professor in the Department of ECE at Karpagam College of Engineering. Developments in biomedical engineering and computer networks keep her fascinated to carry on further research in these domains.