

Advancing Bloodstream Infection Prediction Using Explainable Artificial Intelligence Framework

Rajeev Bopche, MSc^{1*}, Lise Tuset Gustad, PhD^{2,8}, Jan Egil Afset, MD, PhD³, Birgitta Ehrnström, MD, PhD^{4,6,7}, Jan Kristian Damås, MD, PhD^{4,6} and, Øystein Nytrø, PhD^{1,5}

¹ *Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway.*

² *Faculty of Nursing and Health Sciences, Nord University, Levanger, Norway*

³ *Department of Medical Microbiology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*

⁴ *Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway*

⁵ *Department of Computer Science, The Arctic University of Norway, Tromsø, Norway*

⁶ *Department of Infectious Diseases, Clinic of Medicine, St Olavs Hospital, Trondheim, Norway*

⁷ *Clinic of Anaesthesia and Intensive Care, St Olavs Hospital, Trondheim University Hospital, Trondheim*

⁸ *Department of Medicine and Rehabilitation, Levanger Hospital, Nord-Trøndelag Hospital Trust*

Abstract: Bloodstream infections (BSIs) represent a critical public health concern, primarily due to their rapid progression and severe implications such as sepsis and septic shock. This study introduces an innovative Explainable Artificial Intelligence (XAI) framework, leveraging historical electronic health records (EHRs) to enhance BSI prediction. Unlike traditional models that rely heavily on real-time clinical data, our XAI-based approach utilizes a comprehensive dataset incorporating demographic data, laboratory results, and full medical histories from St. Olavs Hospital, Trondheim, Norway, covering 35,591 patients between 2015 and 2020. We developed models to differentiate between high-risk and low-risk BSI cases effectively, optimizing healthcare resource allocation and potentially reducing healthcare costs. Our results demonstrate superior predictive accuracy, particularly the tree-based models, which significantly outperformed contemporary models in both specificity and sensitivity metrics.

Keywords: Machine Learning, eXplainable Artificial Intelligence, Bloodstream infections, Predictive analytics, Electronic Health Records

Author Summary

In this research, we have developed a new tool that uses artificial intelligence to better predict bloodstream infections, which can lead to serious conditions like sepsis if not quickly identified and treated. It is the first of its kind framework that analyzes past health records and helps identify patients who are at high risk of infection more

^{1*} Corresponding Author: Rajeev Bopche, rajeev.bopche@ntnu.no

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

29 accurately than existing tools. Unlike existing tools our framework can be implemented at any stage of the patient
30 trajectory and is the only framework to achieve good accuracy without the use of intimate patient features such as
31 vital signs. This ability could enable doctors to prioritize care more pre-emptively, effectively, potentially saving
32 lives and reducing unnecessary medical tests. Our approach is designed to be easily understood and used by both
33 medical professionals and those with little technical expertise, making it a valuable addition to hospital systems.

34 **1. INTRODUCTION**

35 Bloodstream infections (BSIs) are a significant public health challenge, often leading to severe clinical conditions
36 such as sepsis and septic shock, particularly when unrecognized or untreated. The rapid progression of these
37 infections, coupled with their association with high morbidity, mortality, and healthcare costs, renders BSIs a
38 critical challenge in clinical care[1, 2]. The accuracy of available clinical decision tools for BSI and sepsis is in
39 need of improvement. Currently, most of them are mostly based on changes in vital signs and abnormal blood test
40 results[3, 4]. Improved prediction can lead to more efficient allocation of resources and reduced health costs. A
41 refined initial stratification would aid in allocating resources to patients with a high risk of a BSI and reduce
42 needless testing of patients with low risk. BC may yield not only relevant bacteria causing disease but also growth
43 of contaminating microbes. Ambiguous culture results may be difficult to interpret and lead to clinical uncertainty,
44 often resulting in longer hospital stay and unnecessary administration of antibiotics [5, 6]. Therefore, reduced
45 collection of BC in patients with a low risk of BSI can lead to a higher positive predictive value of blood cultures
46 as well as reduced use of antibiotics. Prompt and early identification of high-risk versus low-risk patients is thus
47 imperative for adequate and effective initial handling of patients with suspected BSI, optimized allocation of
48 healthcare resources and reduction of associated costs.

49
50 The advent of Artificial Intelligence (AI) allows innovative methodologies for BSI diagnostics, showcasing
51 the potential to enhance or even surpass human expertise in diagnostic endeavors [7]. Despite its demonstrated
52 efficacy, the integration of AI into clinical workflows remains limited [8, 9]. Facilitating this integration may
53 involve leveraging AI models to augment guidelines-based Clinical Decision Support Systems (CDSS) rather than
54 striving for the development of fully autonomous AI-based CDSS [10, 11]. Furthermore, adopting a stance
55 grounded in Predictive, Preventive, and Personalized Medicine (PPPM) principles could refine the utilization of
56 AI, emphasizing the analysis of historical rather than real-time data variables [12, 13]. The primary aim of this

57 study was to develop and validate an eXplainable Artificial Intelligence (XAI) framework that utilizes historical
58 electronic health records (EHRs) to enhance the prediction of BSIs, thus facilitating early and precise intervention.
59 By integrating a rich dataset with advanced machine learning techniques, we sought to overcome the limitations
60 of current real-time diagnostic methods. Our findings confirm that the XAI framework not only meets but exceeds
61 traditional diagnostic performance, demonstrating enhanced predictive accuracy and operational efficiency in BSI
62 management. This achievement marks a significant step forward in the application of AI in medical diagnostics,
63 potentially transforming patient outcomes in healthcare settings.

64 **2. LITERATURE REVIEW**

65 BSI are a critical precursor to sepsis, a severe and potentially life-threatening condition. Early detection and
66 management of BSI can significantly mitigate the risk of progressing to sepsis. In this context, leveraging AI
67 models to predict and manage BSI presents a promising avenue. Our literature review sought to capture the breadth
68 and heterogeneity of recent advances in machine learning ML-based models for BSI prediction. We compiled and
69 analyzed thirty studies published in the last five years, focusing on various healthcare settings and patient
70 demographics [14-44]. These studies predominantly focused on inpatient settings, constituting 56% (n=17) of the
71 research, followed by emergency department (ED) settings at 23% (n=7), and ICU settings at 20% (n=6). Within
72 the inpatient group, the studies varied, with nine examining general populations [14-22], two targeting inpatients
73 with central venous catheters (CVC) [23, 24], and others focusing on specific patient categories such as,
74 hemodialysis (HD) patients [26], cancer patients [27], maternity patients [28], patients with low procalcitonin
75 levels ($PCT \leq 2.0$ ng/ml) [29], and HIV patients [30]. Bacteremia was the primary condition under study in 24
76 articles, including varied focuses such as fungemia [14] and Candidemia [27, 43], while three studies aimed at
77 predicting central line-associated bloodstream infections (CLABSIs) [33, 34, 41] and one on hospital-acquired
78 BSI (HA-BSI) [22]. All articles reported high prediction performance (AUROC > 0.7) except for one article [20]
79 which reported poor performance metrics. Most studies were single centered, three articles sourced their data from
80 two hospitals [14, 16, 38], and two articles used data from multiple centres [35, 43]. The key characteristics such
81 as, target condition, number of patients or samples, data source, prevalence, ML models, and top predictors for
82 each study grouped by settings are shown in Table 1 on Page No. 28.

83

84 Among studies with study design of inpatient settings, Bhavani et al. (2020) used EHRs from two academic
85 tertiary medical centers between 2007 and 2018. Data types included Demographic data, International
86 Classification of Diseases billing codes, clinician orders, blood culture results, vital signs, nursing assessments,
87 and laboratory values [14]. The LR and GBM models were developed. GBM models demonstrated superior
88 performance in predicting bacteremia and fungemia with area under the receiver operating characteristic curve
89 (AUC) scores of 0.78 for bacteremia and 0.88 for fungemia. Lee et al. (2019) investigated the early detection of
90 bacteraemia using an artificial neural network (ANN) model, specifically a multi-layer perceptron (MLP). The
91 study utilized data from 13,402 patients at Gangnam Severance Hospital in South Korea, identifying 1,260
92 episodes of bacteraemia from blood cultures [15]. Data included 20 clinical variables such as vital signs and
93 various laboratory data. The study highlighted the effectiveness of MLP models, showing remarkable sensitivity
94 in identifying bacteraemia episodes based on a well-curated set of clinical variables. Lee et al. (2022) expanded
95 on earlier research, utilizing an extensive dataset, also from Gangnam Severance Hospital, consisting of a larger
96 patient sample size and covering more recent patient admissions from 2013 to 2018 [16]. This study employed
97 enhanced ANN models alongside other ML techniques like Random Forest (RF) and Support Vector Machines
98 (SVM) to predict bacteraemia from clinical and laboratory data. The study by Mahmoud et al. (2021) developed
99 a predictive model for bacteremia using data from 7,157 adult patients admitted to King Abdulaziz Medical City
100 in Riyadh. This retrospective cohort study utilized EHRs from July 2017 to July 2019 to analyze 36,405 blood
101 culture tests [20]. The dataset included demographic details, clinical variables such as vital signs (e.g.,
102 temperature, heart rate, blood pressure), and laboratory test results (e.g., white blood cell count, platelet count,
103 creatinine level, lactic acid level, C-reactive protein (CRP), and procalcitonin levels). Several machine learning
104 models were employed to determine the best predictor of positive blood cultures, including Neural Networks
105 (NN), Random Forest (RF), Logistic Regression (LR), Decision Trees (DT), Naive Bayes (NB), and Support
106 Vector Machines (SVM) with a Radial Basis Function (RBF) kernel. The study by Garnice et al. (2021) utilized
107 a dataset from the Hospital Universitario de Fuenlabrada in Madrid, comprising 4,357 patients with 117 features
108 per patient [21]. The features included patient demographics, medical history, clinical analysis, comorbidities, and
109 the results of blood cultures, differentiated into cases of bacteraemia and no bacteraemia. Three supervised
110 machine learning classifiers were implemented: Support Vector Machine (SVM), Random Forest (RF), and K-
111 Nearest Neighbours (KNN). Each classifier was employed in two scenarios: using only the features available at
112 the time of blood extraction, and a second scenario that included additional features revealed during the blood

113 culture process. These studies highlight the potential of ML to enhance diagnostic accuracy and patient outcomes
114 in detecting critical infections.

115

116 Among studies in the ED settings, Choi et al. (2022) developed ML models to predict bacteremia in emergency
117 department (ED) settings at Seoul National University Hospital [31]. The study used data from 24,768 adult
118 patients collected between 2016 and 2018. The models utilized demographic information, chief complaints, vital
119 signs, and laboratory data collected during ED triage and disposition. Two primary models were developed: the
120 Triage XGB model and the Disposition XGB model. In a subsequent study Choi et al. (2023) aimed to refine the
121 predictive accuracy of ED triage-based bacteremia identification using an advanced ensemble of ML techniques.
122 The study analyzed data from over 30,000 ED visits, employing a wide array of clinical inputs including detailed
123 symptom descriptions, vital signs, and initial lab results [32]. The developed model incorporated a Gradient
124 Boosting Machine (GBM) framework that effectively integrated the diverse dataset to predict bacteremia risk.
125 The study by Schinkel et al. (2022) harnessed data from EHRs of 44,123 unique emergency department visits
126 across four hospitals: Amsterdam UMC, Zaan Medical Center, and Beth Israel Deaconess Medical Center,
127 covering the period from 2011 to 2021 [35]. The model employed was a hybrid of LR and (eXtreme Gradient
128 Boosting) XGBoost, with the latter outperforming in predictive accuracy. The data utilized included patient
129 demographics, vital sign measurements, and common laboratory results such as temperature, creatinine, and C-
130 reactive protein, which were identified as the most influential predictors. This predictive model was integrated
131 into the VUMC's EHRs system for real-time prospective evaluation, affirming its practical utility by potentially
132 reducing unnecessary blood culture analyses by at least 30%. The study by Boerman et al. (2022) utilized a single-
133 center, retrospective observational design, the study encompassed data from 51,399 ED visits at the Amsterdam
134 UMC from September 2018 to June 2020 [33]. Data included demographics, vital signs, laboratory and radiology
135 results, and medications administered during ED visits. The study employed two predictive models: a LR model
136 and a gradient boosted tree (GBT) model, both of which demonstrated good predictive performance with an area
137 under the receiver operating characteristic curve (AUROC) of approximately 0.77 and 0.78 respectively. Notably,
138 the gradient boosted tree model was optimized to predict 69% of blood cultures as negative, with a negative
139 predictive value exceeding 94%, indicating its utility in potentially reducing unnecessary blood cultures and
140 associated healthcare costs. The models harnessed a comprehensive array of features, including commonly
141 available clinical data such as C-reactive protein levels and white cell counts, to predict the likelihood of
142 bacteraemia. These studies illustrate how integrating ML models into ED workflows can improve the speed and

143 accuracy of BSI detection, potentially reducing unnecessary interventions and optimizing resource allocation.
144 However, all the model focused on recent data pertaining to the patients and none of the study utilized the
145 predictors from the complete medical history of their patients, apart from demographic and information on co-
146 morbidities. In our previous work through innovative feature engineering from historical medical records and
147 employing an array of machine learning classifiers, we showcased the efficacy of eXtreme Gradient Boosting
148 (XGBoost) model, in predicting 30-day mortality using EHR trajectory features [56].

149 **2. MATERIALS AND METHODS**

150 **2.1. Source of data**

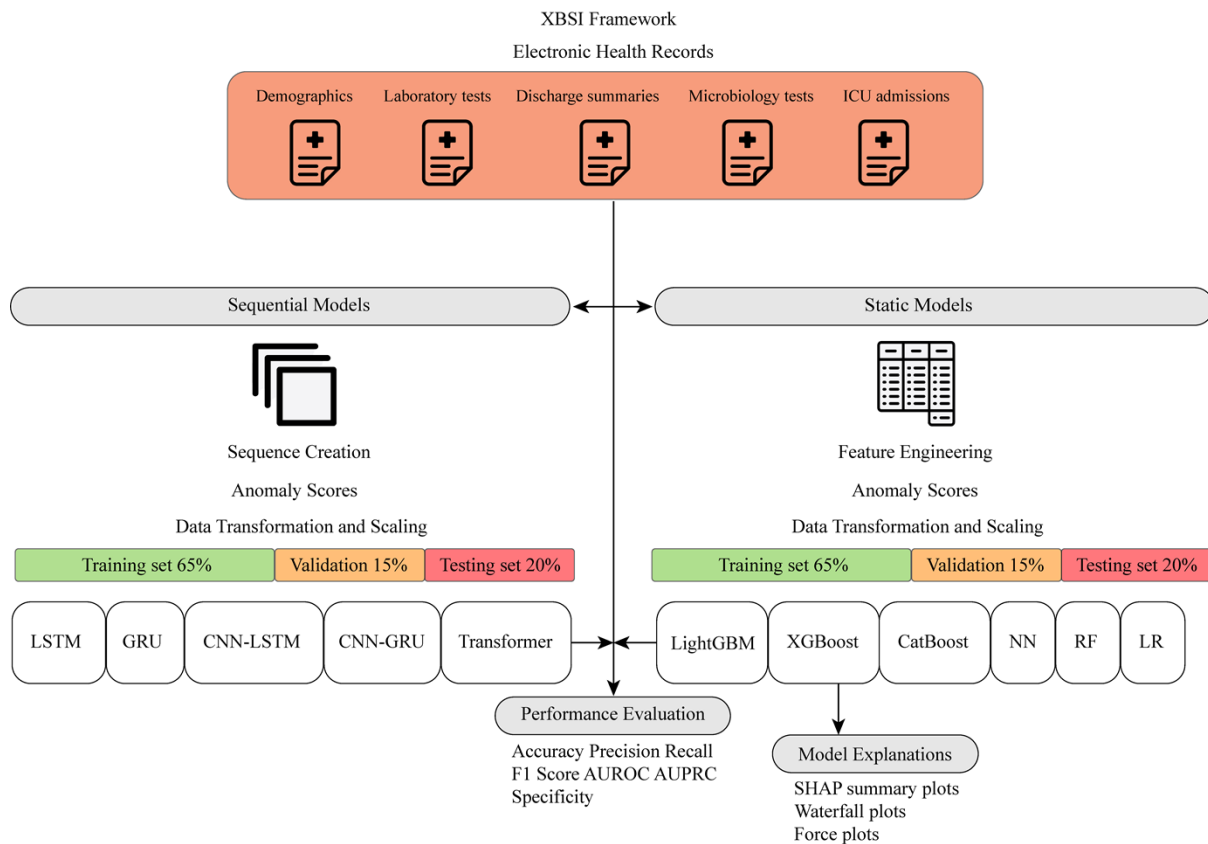
151 EHRs provide a longitudinal perspective of patients' interactions with hospital service. In Norway, with
152 predominantly public specialist healthcare, patients often have long and continuous histories within one hospital's
153 records. This study harnessed EHRs from St. Olavs University Hospital, Trondheim, Norway, encompassing
154 35,591 patients with suspected bloodstream infections (BSIs) identified via physician-initiated blood cultures
155 between 2015 and 2020. The EHRs encompassed, curated data from the inception of electronic records in 1999
156 until 2020, exclusively included hospital care episodes (excluding primary care and other specialist care episodes),
157 Intensive Care Unit (ICU) admission details, microbiology test results, laboratory test results and patient
158 demographics comprising of gender, date of birth, and date of death. Diagnoses and Procedures within these
159 records were classified using the International Classification of Diseases, 10th Revision (ICD-10), facilitating
160 standardized disease identification critical for the analytical models. This study adhered to the 'transparent
161 reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [46]. For ethical
162 considerations the EHRs were de-identified and accessed through a private cloud computing platform.

163

164 **2.2. XBSI Framework**

165 The XBSI framework aggregated various data types from the raw EHRs including demographics, laboratory
166 tests including microbiology tests, discharge summaries, and ICU admissions, as depicted in Figure 1. This dataset
167 underwent preprocessing, event log creation for sequential ML models, and feature engineering for static ML

168 models before being transformed and scaled to facilitate the respective model development pipelines. Following
 169 subsections details of the various steps in the XBSI framework.
 170



171
 172
 173 **Figure 1. Schematic Overview of the XBSI framework.** The diagram illustrates the XBSI framework applied in the study, starting with the
 174 extraction of electronic health records (EHRs), which include demographics, laboratory tests, discharge summaries, microbiology tests, and
 175 ICU stays. The workflow bifurcates into two parallel processes: Sequence Creation and Feature Engineering, both incorporating Anomaly
 176 Scores derived from the data. Subsequent steps include Data Transformation and Scaling, with the data split into training (at 65% and 80%,
 177 respectively), validation (at 15%), and testing (at 20%) subsets. The bottom layer of the workflow depicts the range of ML models, the
 178 performance metrics used to assess them, and the model explanations generated to study the feature importances.

179
 180 **2.2.1. Data Preprocessing and Transformation:**

181 Utilizing Python libraries such as Pandas and NumPy, the CSV files comprising raw EHRs, were loaded into
 182 dataframes for further processing. The discharge summaries required several data cleaning steps to ensure quality
 183 and relevance of the data: Relevant patient information such as identifiers, admission and discharge times, and
 184 diagnostic codes were retained. Instances of missing identifiers were addressed by replacing empty strings with

185 NaN values and subsequently removing these records. Data were organized by patient identifier and
186 admission/discharge times to maintain coherent episode tracking. Non-standard characters within diagnostic
187 codes, such as semicolons and commas, were standardized to spaces, and any duplicates were removed. The
188 timestamps were converted into datetime format, facilitating the calculation of the length of stay in hours for each
189 episode. The duration of each ICU stay was calculated in hours, along with the total count of each type of hospital
190 admissions per patient. The request dates in both laboratory test results and microbiology test results were
191 standardized to datetime objects and used to create eventlogs of tests per patient. Specialized functions were
192 created to clean the laboratory and microbiology test table entries. This function performed tasks to remove any
193 non-numeric characters, which could represent encoding errors or artifacts from data entry. It standardized
194 decimal point characters by replacing commas with periods, which is necessary for consistent numerical
195 representation across different regions that may use varying formats for decimal points. The microbiology test
196 table was filtered to identify suspected BSI episodes, with an emphasis on blood culture tests. The results column
197 was processed to standardize and clean the values, categorizing them as 'positive', 'negative', or 'contaminant'
198 based on the results column. For each patient ID, four event logs were created, from discharge summaries, ICU
199 admissions, laboratory tests and microbiology tests, following subsection describes the event logs.

200 2.2.2. Event log description:

201 *Discharge Summaries Event Log:* Captures patient discharge information, including admission and discharge
202 times, diagnostic and procedural codes, urgency and care level code.

203 *ICU Admissions Event Log:* Records details of each ICU stay, including the duration in hours and the total count
204 of ICU admissions per patient and total length of ICU stays per patient.

205 *Laboratory Tests Event Log:* Includes results of various laboratory tests standardized and organized
206 chronologically for each patient.

207 *Microbiology Tests Event Log:* Consists of microbiology test results, grouped by collection sample type, and
208 categorizes them based on outcomes such as 'positive', 'negative', or 'contaminant'. Groups of microbiology
209 tests categorized by collection sample type are given in Supplementary List 1.

210 2.2.3. Sequence creation:

211 The sequence creation process was implemented using the `create_sequences` function, designed to compile a
212 comprehensive view of a patient's medical history over their entire recorded history. The `create_sequences`

213 function systematically constructs a timeline of medical events for each patient, for each BC test. The medical
214 events were merged from the four event logs. The filtered event logs are merged into a single dataframe, ensuring
215 no information is lost. This step involves an outer join on patient ID and date, maintaining all records from each
216 event log.

217 2.2.4. Feature engineering:

218 This approach involved creating a dataset with attributes derived from ICD codes, procedure codes, laboratory,
219 and microbiology test results. Laboratory test results were organized using pivot tables, ensuring a structured
220 format for analysis. Tests such as 'bilirubin' (total, conjugated, and unconjugated), 'C-Reactive Protein' (CRP),
221 and 'lactate' (various measurements) were included, alongside white blood cell count (leukocytes), platelet count
222 (thrombocytes), and blood gas measurements (pH, PO₂). Similarly, microbiology test results were consolidated
223 to reflect various sample types such as blood, urine, and other fluids, employing a dictionary mapping to
224 streamline similar types. The resulting pivot table included columns for diverse samples, ranging from 'blood
225 culture test' to 'urine', 'feces', and 'nasal swabs'. A function was created to determine the aggregated result of blood
226 culture tests for each patient group, considering the possibility of concurrent positive and contaminant results.
227 This aggregation provided a comprehensive view of the infection status per patient. Moreover, the history of prior
228 positive, negative, and contaminant results was calculated and added to the dataset, offering a valuable perspective
229 on the patient's previous encounters with BSIs. Comorbidities were extracted and processed to identify unique
230 diseases from patient records. New columns were created for each disease, and the counts were updated based on
231 patient history. Finally, aggregate columns were added for both ICD and procedure codes to calculate cumulative
232 sums. This method enabled the capturing of the cumulative history of medical conditions and procedures for each
233 patient. All the NaN values representing absence of a condition or measurement were filled with zero and "0" is
234 not interpreted as a value.

235 2.2.5 Model Development

236 This section outlines the development and evaluation of ML models. The process bifurcates into the sequence
237 creation and feature engineering pipelines for the temporal and static ML pipelines respectively. An Isolation
238 Forest model was implemented to detect anomalies within the data. This model was trained exclusively on normal
239 data (negative BC episodes) and then used to compute anomaly scores for the training, validation, and test sets.
240 These scores were normalized and appended to the original dataset to serve as additional features, enhancing the

241 model's ability to distinguish between normal and anomalous patterns. Given the sequential and tabular nature of
242 the input data for the respective ML pipelines, the necessary transformation steps were employed to prepare it for
243 the corresponding learning algorithms. The data was reshaped and scaled using the StandardScaler from
244 sklearn.preprocessing. This normalization step is crucial for models that are sensitive to the scale of input features.
245 After scaling, the data was reshaped back to its original form, ensuring compatibility with the machine learning
246 models used. Temporal ML models included Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU),
247 Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM), Convolutional Neural Network-Gated
248 Recurrent Unit (CNN-GRU), and Transformer models. These models were trained on sequences of medical events
249 to capture temporal dependencies within the data. On the other hand, the static ML models, which treat data points
250 as independent and identically distributed, comprised of Light Gradient Boosting Machine (LightGBM), eXtreme
251 Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), Artificial Neural Network (ANN), Random
252 Forest (RF), and Logistic Regression (LR). Detailed information of each ML model is given in Supplementary
253 Methods, Section 1.1., 1.2.

254 2.4.6 Model Validation

255 For cross validation, we chose a 5-fold cross-validation scheme. This technique involves dividing the entire dataset
256 into 'k' equally sized subsets or folds. The model is then trained on 'k-1' folds and tested on the remaining fold.
257 This process is repeated 'k' times, each time with a different fold as the validation set. In addition to cross-
258 validation, we also validated our model on an independent test set, which was separated from the dataset at the
259 outset and not used during the training phase. The test set comprised of the most recent 20% of the BC episodes
260 while the earliest 80% of BC episodes were assigned as training cohort. Further, the training data was divided into
261 a smaller training set and a validation set, with 15% of the training data allocated for validation. This additional
262 split allowed for the tuning of hyperparameters and the assessment of the model's performance during training.
263 Each model was evaluated using a variety of metrics to evaluate model performance, including accuracy,
264 precision, recall, F1-score, specificity, the Area Under Precision-Recall (PR) Curve (AUPRC) and the Area Under
265 the Receiver Operating Characteristic (AUROC). Detailed information on each performance metric is given in
266 Supplementary Methods, Section 1.3.

267 2.4.7 Model Explanations

268 For interpreting the outputs of our ML models, we employ both global and local explanation methods to provide
269 comprehensive insights into feature importances which guide clinical decision-making. To evaluate the overall
270 impact of features across all predictions, we generate SHAP (SHapley Additive exPlanations) summary plots [55].
271 These plots aggregate SHAP values to illustrate the average influence of each feature on the model output, ranked
272 by significance. This method, rooted in game theory, decomposes a prediction into the contribution of each
273 feature, providing a transparent view into the predictive process. SHAP's suitability for tree-based models was
274 enhanced by its efficient computation of exact SHAP values using the TreeExplainer algorithm [57], which
275 significantly reduced computational complexity by exploiting the structural properties of decision trees [58]. By
276 integrating SHAP summary plots, we can discern which features are most instrumental globally in predicting
277 bloodstream infections, facilitating an understanding of model behavior over the entire dataset

278

279 On an individual prediction level, we utilize waterfall plots and force plots to detail the contribution of each feature
280 to specific predictive outcomes. Waterfall plots provide a step-by-step breakdown of how each feature's value
281 contributes to the final prediction, starting from the base value (the average model output across all data points)
282 and adding the effect of each feature sequentially. This visualization helps in understanding the decision-making
283 process for individual predictions, which is crucial for clinical validation and personalized patient insights. Force
284 plots, another type of local interpretability tool, display how each feature's value pushes the model's prediction
285 higher or lower, which is particularly useful for individual patient assessments. These plots highlight the positive
286 or negative contribution of each feature towards the final prediction, allowing healthcare providers to grasp the
287 underlying reasons for a model's decision on a case-by-case basis.

288 **2.2. Participants and Outcome**

289 All adults (aged ≥ 18 years) who had at least one BC episode during their hospital stay or visit, which was
290 identified and ordered by a physician on the grounds of suspicion of a BSI. The primary outcome was whether a
291 blood culture episode was positive or negative for bacteremia. A BC episode was defined as a distinct
292 nonoverlapping 24-hour period in which one or more BC tests were ordered [44]. If one or more results within a
293 BC episode were positive, then the BC episode was considered positive. BC results with contaminants were

294 considered negative results [45]. The list of microbes considered as contaminants is given in Supplementary Table
295 6.

296 **2.3. Predictors**

297 We used data available till the date of a BC episode for training the ML prediction models. The predictors included
298 age, sex, results of the recent laboratory test values, previous positive microbiology tests, count of co-morbidities,
299 ICD and procedural codes, and total ICU stays. The most common laboratory tests were Bilirubin, C-Reactive
300 Protein (CRP), Creatinine, Leukocytes, and Thrombocytes. The counts of prior positive results of microbiology
301 tests grouped by their collected sample type were calculated and used as predictors of previous history of
302 infections (Supplementary List 1). From the medical history the predictors included, counts of the occurrences of
303 different ICD-10 diagnostic and the Nordic Medico-Statistical Committee (NOMESCO) Classification of Surgical
304 Procedures (NCSP), codes and Classification of Medical Procedures (NCMP) codes, were classified according to
305 the initial character (alphabetic), corresponding to the various chapters in of the ICD-10. Each patient record was
306 expanded with new columns for the counts in the recent and complete history of the corresponding to each
307 character, incrementing the count for each instance where a character led the code. For calculation of diagnostic
308 and procedural codes counts in the recent episode, the current episode and any admissions or visits within one
309 month of the BC test date were merged into the current medical episode. The LOS feature stored the total length
310 of stay in the hospital for the current episode till the time of BC. The total LOS feature stored the value of the
311 cumulative hospital length of stay (LOS) per patients. The description of each predictor and its mean across the
312 dataset is given in the Supplementary Table 4. Further stratification was conducted by categorizing ICD codes
313 into disease groups pertinent to clinical significance, such as 'explicit sepsis', 'infection', and 'organ dysfunction',
314 among others. The table depicting the ICD codes selected for different disease groups is given in Supplementary
315 Table 5.

316 **3.4. Prediction task modeling**

317 We undertook prediction modeling using two distinct datasets: a static model dataset, X , and a sequential model
318 dataset, Y . The static dataset, X , encompasses labelled aggregated, patient-specific information available till the
319 day of BC episode and labels for. Using X , we trained static models M_{sta} to predict the likelihood of a positive
320 BC episode. the sequential dataset, Y , contains labelled sequences of medical events per patient compiled from

321 the eventlogs for each BC test. The sequential models M_{seq} is developed using Y to predict the likelihood of
322 positive BC episode.

323

324 **2.5. Statistical Analysis**

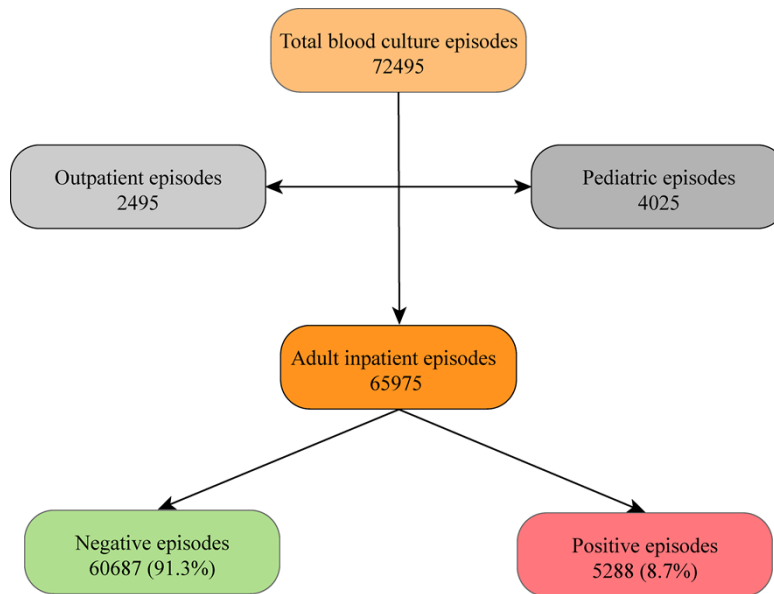
325 The mean values for the significant features across the two classes were computed, providing an understanding
326 of how each feature varies with the BC episode results. Statistical analysis was performed using independent t-
327 tests to compare the means of each feature between two independent groups labeled by blood culture test results
328 [47]. A Pearson correlation matrix was constructed for the significant features to examine the strength and
329 directionality of the relationships between them [48]. The resulting coefficients was visualized using a heatmap.

330 **3. RESULTS**

331 **3.1. Patient Characteristics**

332 There were a total of 72,495 BC episodes in the dataset. Following the exclusion of pediatrics and outpatient BC
333 episodes 65,975 adult inpatient BC episodes were included in the analysis. Of the BC episodes, 5,288 (8%) were
334 classified as positives. Please see the flow chart provided in Figure 2. The differences in the mean values for all
335 feature across the two classes along with the T-statistic and p-value is given in Table (Supplementary Results,
336 Supplementary Table 2). The top 25 most significant features and the comparison between their mean values
337 across the two classes, along with the T-statistic and p-value are given as Table 1. There were significant
338 differences between the positive and negative blood culture groups regarding the occurrence of ICD codes starting
339 with A, B, and N in the medical records. The ICD Chapter I (A00-B99) describes certain infectious diseases and
340 parasitic diseases and ICD Chapter XIV (N00-N99) concerns kidney and urinary disorders. From laboratory
341 markers, bilirubin, creatinine, and CRP showed the most significant differences between the groups. The positive
342 BC group also had higher mean age, higher total length of stay (LOS) till the time of BC, shorter time to the last,
343 most recent episode, and a higher number of previous positive BC test results. The correlation matrix heatmap
344 depicts a visual representation of the strength and directionality of correlations between various clinical features
345 within the EPRs, highlighting the complex interdependencies relevant to BSI prediction, and is presented in
346 Supplementary Figure 5. The top three most correlated coefficients among significant features is given in

347 Supplementary Table 3. The list of all the features, their description, and the mean value across the dataset is given
 348 in Supplementary Table 4.



349

350

Figure 2. Flow chart The flowchart depicts the categorization of the BC episodes.

351

Comparison of the predictors (Top 25 most influential) between the two classes

Feature names	Class 1 (mean) (n=5288)	Class 0 (mean) (n=60687)	T-Statistics	P-Value
ICD_A (Infectious diseases)	0.344175492	0.127045331	-41.35067292	0
ICD_B (Infectious diseases)	0.270990923	0.098044062	-34.86262822	6.9733E-264
ICD_N (Kidney disease)	0.498108926	0.255112298	-31.06397543	2.4511E-210
BILIRUBIN TOTAL (Total bilirubin)	14.64849131	7.316773443	-23.83455936	6.0583E-125
prior_positive_blood_culture_test	0.370083207	0.178852143	-20.33423563	1.22442E-91
ICD_J (Influenza, pneumonia etc)	0.307677761	0.488885593	16.95136079	2.5746E-64
ICD_R (abnormal symptoms and findings)	0.298411498	0.188771895	-16.06835215	5.48158E-58

Age (at the time of BC)	68.87821483	65.07332707	-14.51708417	1.11892E-47
KREATININE (Creatinine)	99.42019885	82.73778714	-13.30695014	2.42534E-40
CRP	110.8463237	94.62243471	-12.20533489	3.21434E-34
ICD_A_aggregate	0.797655068	0.557236311	-11.29534135	1.47299E-29
LEUKOCYTTER (Leukocytes)	11.57262042	10.27756163	-10.07214858	7.68314E-24
procedure_G_aggregate	0.996596067	1.48710597	9.453454788	3.3812E-21
ICD_U (Unsure etiology and antimicrobial resistance)	0.018910741	0.006953713	-9.34670256	9.3135E-21
ICD_J_aggregate	1.876134644	2.99597937	9.271600994	1.88697E-20
procedure_K (Procedures on the ears)	0.073940998	0.038377247	-8.468324845	2.5398E-17
TROMBOCYTTER (Thrombocytes)	156.5213799	174.202004	8.409629616	4.20742E-17
ICD_K (Gastrointestinal diseases)	0.234871407	0.179527741	-8.148452108	3.75005E-16
procedure_J (Procedures on endocrine glands)	0.184379728	0.127407847	-6.851217934	7.38635E-12
prior_positive_edta	0.142965204	0.254074184	6.204949673	5.50384E-10
ICD_M (Musculoskeletal disease)	0.067133132	0.092490978	5.438316649	5.39768E-08
ICD_B_aggregate	0.752458396	0.554468008	-5.194454856	2.06E-07
ICD_R_aggregate	2.291603631	2.686555605	5.06204369	4.15899E-07
Explicitsepsis (Sepsis episdoe)	0.006807867	0.002751825	-5.034421936	4.80552E-07
LOS (Length of stay till BC tests)	8.751426185	8.013890262	-4.578791274	4.68523E-06

352

353

Table 1. Comparison of the predictors (Top 25 most influential) between the two classes.

354

355 3.2. Model Performance

356 Comparative performance metrics of the ML Models is given as Table 2. The Prediction performance for the ML

357 models was evaluated using the testing set. For the sequential models, the CNN-LSTM model achieved the highest

358 AUC of 0.7182 (95% CI 0.7064-0.7300) and AUPRC of 0.2752. The LSTM model had the highest Accuracy of

359 0.8955 and Precision of 0.3084. The Transformer model demonstrated the highest Recall/sensitivity of 0.4688.

360 The GRU model gave the highest F1-score of 0.3349. In contrast, the static models outperformed the sequential
 361 models, with the CatBoost model exhibiting the highest AUC of 0.8167 (95% CI 0.8049-0.8285) and F1-score of
 362 0.4132. The RF model showcased the highest AUC of 0.8410 (95% CI 0.8292-0.8528). However, it is notable
 363 that both the RF and NN models, with their high accuracy and specificity scored least in terms of recall. On the
 364 other hand, the LR model achieved the highest recall, in contrast to the lower performance across other metrics.
 365 The balanced performance of both XGBoost and LightGBM models across all metrics underscores the robustness
 366 of tree-based models compared to their sequential counterparts.
 367

Model	AUC (95% CI)	Accuracy	Precision	Recall	F1 Score	AUPRC	Specificity
<i>Sequential Models</i>							
LSTM	0.6961 (0.6843-0.7079)	0.8955	0.3084	0.2767	0.2913	0.2503	0.9422
GRU	0.7155 (0.7037-0.7273)	0.8652	0.2862	0.4063	0.3349	0.2599	0.8985
CNN-LSTM	0.7182 (0.7064-0.7300)	0.8826	0.3048	0.3529	0.3265	0.2752	0.9224
CNN-GRU	0.7020 (0.6902-0.7138)	0.8697	0.2735	0.3475	0.3054	0.2546	0.9093
Transformer	0.7015 (0.6897-0.7133)	0.8119	0.2829	0.4688	0.3047	0.2307	0.8351
<i>Static Models</i>							
XGBoost	0.7951 (0.7833-0.8069)	0.8591	0.3533	0.5680	0.4025	0.3942	0.8876
LightGBM	0.8105 (0.7987-0.8223)	0.8103	0.3441	0.6628	0.3737	0.3987	0.8259
CatBoost	0.8167 (0.8049-0.8285)	0.8529	0.3523	0.6274	0.4132	0.4203	0.8750
NN	0.7693 (0.7575-0.7811)	0.9261	0.5126	0.2886	0.3634	0.3618	0.9775
LR	0.7751 (0.7633-0.7869)	0.7534	0.2462	0.6812	0.3219	0.2861	0.7627
RF	0.8410 (0.8292-0.8528)	0.9218	0.8510	0.2201	0.1990	0.4545	0.9882

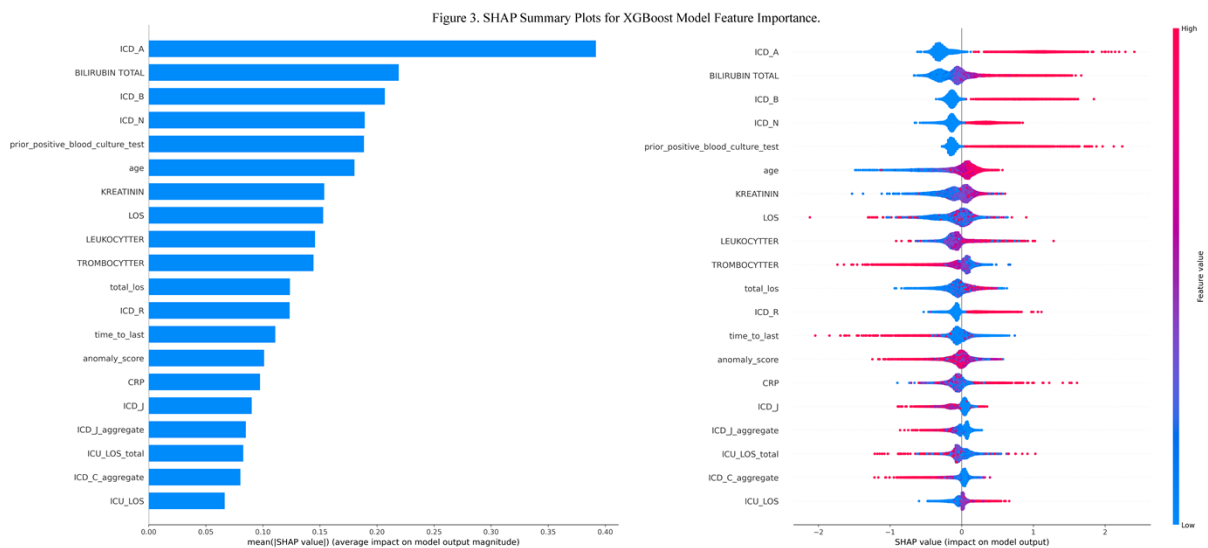
368

369

Table 2. Comparative Performance Metrics of the ML Models.

370 3.3. Global Feature Importances

371 The SHAP summary plots in Figure 3. reveal the feature importances derived for the XGBoost model. The bar
372 plot indicates that the count of ICD codes starting with letters A, B, and N, bilirubin (BILIRUBIN TOTAL),
373 creatinine (KREATININ), leukocytes (LEUKOCYTTER), thrombocytes (TROMBOCYTTER), age, number of
374 previous positive blood culture results, and the LOS of current or recent hospital episode were the top ten most
375 important features for predicting positive blood culture. The beeswarm plot further elucidates the directional
376 impact of each feature's value on the model's predictions. Higher values of bilirubin, creatinine, and leukocytes
377 (as indicated by red dots on the right side of the x-axis) lead to a greater increase in the model's prediction values,
378 potentially signifying a higher risk whereas a lower value for thrombocytes signifies a higher risk of bloodstream
379 infection. Other significant factors towards a positive prediction were higher cumulative hospital length of stay
380 (total_los), lower time to the recent episodes, and higher CRP.



381
382
383 **Figure 3. SHAP Summary Plots for XGBoost Model Feature Importance.** Bar plot on the left illustrating the global feature importance
384 ranked by the sum of SHAP values across all samples. On the right is the Beeswarm plot detailing the individual SHAP values for each
385 feature and their impact on the model's output.

386 3.4. Local Feature Importances

387 The figure 4. on Page No. 32. Details the waterfall and force plots for the first three prediction tasks. In the first
388 prediction task, the feature ICD_A (representing infectious and parasitic diseases) has a significantly high positive
389 SHAP value, indicating that an increase in counts of ICD-10 Chapter I codes strongly sways the model towards

390 predicting a bloodstream infection. The 'anomaly_score' feature contributes negatively, suggesting that anomalies
391 detected by the model are not predictive of the outcome. Similarly, creatinine level and ICD_N (genitourinary
392 diseases) show negative SHAP values, implying that higher levels of creatinine and the presence of genitourinary
393 diseases are linked to a lower probability of bloodstream infection in this instance. For the second prediction,
394 counts of prior positive BC results has the most substantial positive impact, which aligns with clinical reasoning
395 that past positive tests could indicate a higher risk for current infection. Additionally, ICD_B (infectious and
396 parasitic diseases) and ICD_R has a positive effect, and total bilirubin has a positive contribution, suggesting high
397 risk of BSI. In the third prediction, similar to the second prediction, counts of prior positive BC results again
398 shows a positive influence. Additionally, 'ICD_C_aggregate' (aggregate count of cancer-related codes) exerts a
399 negative influence on the prediction outcome, whereas 'procedure_W_aggregate' (count of procedures on female
400 reproductive organs in the history) is associated with an increase in the risk of BSI.

401 **4. DISCUSSIONS**

402 In this study we presented the XBSI framework for early prediction of BSI in the hospitals. For the sequential
403 models, the AUC ranged from 0.6961 to 0.7182, indicating a modest discriminatory capacity, with CNN-LSTM
404 marginally leading in performance. Notably, the Transformer model demonstrated the highest sensitivity 0.4688.
405 In contrast, static models depicted better AUC values ranging from 0.7751 to 0.8410. Our findings underscore the
406 superior performance of static ML models, particularly those employing tree-based algorithms such as CatBoost,
407 which achieved a notable AUC of 0.8167 [52]. The statistical analysis and global and local SHAP value
408 interpretations reveal that the count of diagnostic and procedural ICD codes and routine laboratory significantly
409 contribute to the model's predictions. These findings align with existing literature, underscoring the clinical
410 relevance of laboratory tests and medical history for early BSI prediction [14, 15, 16, 20, 21, 51]. The
411 demonstrated efficacy of our XBSI framework highlights the predictive capability of combining medical history
412 with routine laboratory tests for early BSI detection. The modular design of our framework reinforces robustness
413 to all types of medical data and emphasize the utility of integrating diagnostic and procedural ICD codes into
414 predictive models. This approach allows for a nuanced understanding of patient profiles, which is pivotal for
415 implementing predictive healthcare solutions effectively.
416

417 Comparing the recent studies on predicting BSI in hospital settings, Bhavani et al. (2020) utilized LR and GBM
418 models to predict bacteremia and fungemia from EHRs, achieving AUCs of 0.73 and 0.88 respectively. The key
419 predictors identified in the study were time from admission to BC, temperature, age, heart rate, prior
420 bacteremia/fungemia, white blood cells (WBC), blood urea nitrogen (BUN), glucose, diastolic blood pressure
421 (DBP), and systolic blood pressure (SBP) [14]. Lee et al. (2019) compared multiple ML algorithms, including
422 MLP, SVM, and RF, achieving their best AUC of 0.732 with the RF model [15]. Our study surpasses this
423 benchmark, suggesting that the sequential analysis of patient data may capture the temporal dynamics of BSI
424 development more effectively than models relying solely on static point-in-time data. The key predictors reported
425 by the study included, alkaline phosphate (ALP), platelet, maximum body temperature, SBP, WBC, CRP, ICU
426 stay, hospital day to blood culture, age, heart rate, prothrombin time, and albumin. In a subsequent study, Lee et
427 al. (2022) further explored MLP, RF, and XGBoost among others, focusing on their application over a long-term
428 dataset. Their models achieved AUCs of 0.762 for MLP and 0.758 for RF in the 12-hour data group, which are
429 slightly lower than our CatBoost model's performance. the key predictors in this study were monocyte, platelet,
430 hospital stay, neutrophil, total bilirubin, BUN, albumin, ALP, WBC, CRP, creatinine, pulse rate, and chloride
431 [16]. The study by Mahmoud et al. used data from a tertiary care center comprising patient demographics, length
432 of stay before blood culture collection, presence of central line, vital signs, laboratory results, and SIRS and
433 qSOFA scores [20]. They employed various models including NN and LR, with their best models achieving
434 modest performance metrics (highest specificity at 89% but with low sensitivity). This study extensively used
435 vital signs and other real-time clinical parameters such as temperature and heart rate as predictors. Despite this,
436 the highest sensitivity achieved was only 31% with Logistic Regression, and even though some of their models
437 achieved high specificity, they struggled with very low sensitivity, limiting their practical utility in clinical
438 settings. Our model's ability to outperform these studies without the need for immediate clinical data or vital signs
439 not only showcases the robustness and efficiency of our approach but also suggests a potential for earlier and
440 simpler implementation in clinical workflows. This is particularly advantageous in healthcare settings where
441 immediate comprehensive data collection is challenging, offering a powerful tool for early BSI detection that is
442 less dependent on the specific timing of clinical data acquisition. Similar to our study, the study by Garnica et al.
443 (2021) used SVM, RF, and K-nearest neighbors (KNN), with a combination of RF and SVM yielding the most
444 robust performance metrics [21]. The number of days in ICU before BC extraction, presence of catheters, age,
445 chronic respiratory disease, fever, and CRP were reported as key predictors. Their better performance in

446 comparison can also be attributed to the fact that they used a BSI dataset with prevalence rate of 51.3% which is
447 highest reported among all the studies reporting ML based BSI prediction models.

448

449 Our study's findings also contribute to the evolving landscape of ML applications for BSI prediction in ED
450 settings. Schinkel et al. employed LR and XGBoost models using vital signs, laboratory results, and demographics
451 [35]. Their XGBoost model achieved an AUROC of 0.81 (95% CI 0.78-0.83), slightly higher compared to our
452 XGBoost model AUROC of 0.7951 (95% CI 0.7833-0.8069) and an AUPRC of 0.34 (95% CI 0.29-0.38) slightly
453 lower than our XGBoost model AUPRC of 0.3942. The key predictors brought out by this study were temperature,
454 creatinine, CRP, lymphocytes, DBP, bilirubin, thrombocytes, neutrophils, APL, heart rate, SBP, leukocytes,
455 glucose, age, potassium, BUN, sodium, monocytes. Their model effectively reduced unnecessary blood cultures
456 by approximately 30% during real-time prospective evaluation, which aligns with our goals of enhancing
457 diagnostic efficiency and reducing healthcare costs. Our study builds on this foundation by implementing a similar
458 ML approach but extends its application by incorporating a wider array of clinical variables derived from historical
459 EHRs and employing a novel algorithmic configuration that may provide improved predictive performance.
460 Similar to our study, Boerman et al. focused on the ED setting of a large teaching hospital, developing predictive
461 models specifically for BSI outcomes based on data available at the end of ED visits, such as demographics, vital
462 signs, administered medications, and laboratory results. They reported an AUC of 0.77 for their gradient-boosted
463 tree (GBT) model and 0.78 for their LR model, indicating good performance in predicting bacteremia in ED [33].
464 The key predictors in their models were bilirubin, urea, lymphocytes, pulse rate, CRP, neutrophil, age,
465 Temperature, DBP, potassium, glucose, thrombocytes, creatinine, ALP, SBP, and organ damage. The study
466 highlighted the ability of their models to significantly reduce unnecessary blood cultures by predicting negative
467 outcomes with a high degree of accuracy, reflected by a negative predictive value of over 94%. In parallel, our
468 model not only confirms these findings but also demonstrates improved predictive accuracy and efficiency in
469 differentiating between positive and negative BSI outcomes, which could further optimize the use of resources in
470 ED settings. In another study, Choi et al. (2022) demonstrated the effectiveness of XGBoost model, to predict
471 bacteremia at different stages of patient care in the ED, achieving an AUROC up to 0.853 [31]. The key predictors
472 in the study were chief complaint, age, temperature, heart rate, and DBP at triage stage, and neutrophils, platelets,
473 CRP, chief complaint, and creatinine at disposition stage. Their phased approach using predictions at both triage
474 and disposition stages aligns with our methodology of employing dynamic modeling to adapt predictions based
475 on real-time data updates. However, our model differentiates itself by utilizing a more complex array of inputs

476 and a refined computational technique that enhances predictive generalizability. By extending the predictive
477 capabilities beyond structured data, our study underscores the importance of comprehensive data utilization in
478 developing models that can more effectively guide early clinical decisions in real-time settings. Our model's ability
479 to accommodate a broader array of clinical variables may explain any improvements in prediction robustness
480 compared to the framework used by Choi et. al. (2023) [32]. The key predictors in their study were age, vital
481 signs, history of chills, ambulance use. The collective insights from these comparisons suggest that while our
482 model shares common ground with existing approaches, it also explores additional layers of complexity such as
483 patient recent hospital interactions, comorbidities, and previous history of infections, which may influence the
484 generalizability and effectiveness of the model across diverse healthcare environments, capturing more subtle
485 nuances of BSI risk factors and could be pivotal in reducing BSI misdiagnosis.

486

487 Our framework's modular design facilitates the inclusion of additional data types as they become available,
488 enhancing its adaptability across various clinical settings. This feature is particularly valuable in current healthcare
489 environments where flexibility and comprehensive data utilization are crucial for advancing diagnostic accuracies.
490 Moreover, the intuitive nature of the XBSI framework ensures that it can be seamlessly integrated into existing
491 clinical workflows, making it a practical tool for clinicians seeking to leverage AI for improved patient outcomes.
492 The principal clinical value of our approach lies in the ability to identify patients at low risk of a positive BC, at
493 the time of suspicion of BSI without the need for waiting to capture vital signs at the moment or within a specified
494 time window, which could increase patient risk and stay [53, 54]. Integrating our proposed framework in hospitals
495 as a pre-emptive BSI prediction tool can reduce BC ordering and its resulting costs and harms [49]. As it is
496 reported in the literature that the use of data that is not routinely captured in clinical practice is the main reason
497 why none of the prediction models have been implemented in the clinical practice yet [50]. Moreover, in the
498 context of ICU patient monitoring, the application of AI to real-time data may seem redundant. Patients in critical
499 care are already under intensive surveillance, and the utilization of AI for immediate alert systems could
500 potentially clutter the workflow rather than enhance it. Instead, a more strategic use of AI lies in its ability to
501 predict a patient's worsening condition well before critical thresholds are reached. By analyzing historical EHRs,
502 AI can identify subtle patterns and indicators of decline that precede acute episodes, thereby enabling preemptive
503 medical interventions. Moreover, as the complexity of predicting health events increases the earlier the prediction
504 occurs, achieving perfect accuracy, as often highlighted through metrics like AUC in ICU settings [38-43], may
505 not be as critical as maintaining reasonable predictability at the initial stages of patient contact. This approach

506 could shift the focus from crisis response to proactive patient care management, optimizing outcomes through
507 early and targeted intervention. The only limitation of this study is its reliance on data from a single center, which
508 may not represent the diverse patient demographics. Despite these constraints, our model's ability to integrate a
509 broader spectrum of data and apply a real-time learning framework supports its potential utility in clinical settings,
510 promising reductions in unnecessary interventions and improvements in patient management. Furthermore, our
511 use of historical patient data and variety of ML techniques allowed for a more flexible and scalable prediction
512 framework. This aspect of our model enhances its applicability and provides a significant contribution towards
513 advancing BSI prediction using AI.

514 **5. CONCLUSIONS**

515 This study presents a simple and intuitive XAI framework which comprehensively captures the complete medical
516 history of a patient to accurately predict the risk of BSIs. The XBSI framework is the first of its kind to
517 significantly enhance the predictive analysis by integrating information stored as diagnostic and procedural ICD-
518 10 codes. This novel approach diverges significantly from traditional real-time monitoring systems, emphasizing
519 incorporating comprehensive historical patient data. The key predictors include the count of ICD-10 codes for
520 infectious diseases, and kidney and urinary disorders, total bilirubin, creatinine, leukocytes, thrombocytes, age,
521 prior positive blood culture tests, and the total length of hospital stay till the blood culture test. To further enhance
522 the generalizability of our findings and adoption of our framework, there is a need to validate our framework on
523 administrative datasets of hospitals outside Norway and include more diverse data sources, such as biomarkers,
524 genotypes, and phenotypes in future works.

525

526 **Acknowledgements**

527

528 We would like to thank the researchers at Mid-Norway Centre for Sepsis Research for valuable discussions and
529 feedback.

530

531 **Author's contributions**

532 RB conceptualized and designed the study, with inputs from ØN, JEA, BE and JKD. ØN and LGT secured the
533 data and funding for this project. RB performed data preprocessing, model development and interpretation. RB

534 wrote the initial draft of the paper, to which the remaining authors provided comments and feedback. All authors
535 reviewed and approved the final manuscript.

536

537 **Availability of data and materials**

538 The jupyter notebooks implementing the data preprocessing, sequence creation, feature engineering and model
539 development pipeline are available through the corresponding author. The processed and transformed final
540 datasets, and the list of derived features are available through the corresponding author on a reasonable request.

541

542 **Ethics**

543 The use of the EPRs data in this project has been approved by the Regional Committees for Medical and Health
544 Research Ethics (REK) in Central Norway by REK no. 2020/26184.

545

546 **Conflict of interests**

547 None declared.

548

549 **Funding**

550 Financial support for this study was provided by the Computational Sepsis Mining and Modelling project
551 through the Norwegian University of Science and Technology Health Strategic Area.

552

553

554 **References**

555

556 [1] Goto M, Al-Hasan MN. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe.

557 Clin Microbiol Infect. 2013;19(6):501-509. doi: 10.1111/1469-0691.12195.

558 [2] Coburn B, Morris AM, Tomlinson G, Detsky AS. Does this adult patient with suspected bacteremia require blood cultures? JAMA.

559 2012;308(5):502-511. doi: 10.1001/jama.2012.8262.

560 [3] Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized

561 Patients. JAMA Intern Med. 2021;181(8):1065-1070. doi: 10.1001/jamainternmed.2021.2626.

562 [4] Schootman M, Wiskow C, Loux T, Meyer L, Powell S, Gandhi A, Lacasse A. Evaluation of the effectiveness of an automated sepsis

563 predictive tool on patient outcomes. J Crit Care. 2022;71:154061. doi: 10.1016/j.jcrc.2022.154061.

- 564 [5] Bates DW, Cook EF, Goldman L, Lee TH. Predicting bacteremia in hospitalized patients. A prospectively validated model. *Ann Intern*
565 *Med.* 1990;113(7):495-500. doi: 10.7326/0003-4819-113-7-495.
- 566 [6] Sakarikou C, Altieri A, Bossa MC, Minelli S, Dolfa C, Piperno M, Favalli C. Rapid and cost-effective identification and antimicrobial
567 susceptibility testing in patients with Gram-negative bacteremia directly from blood-culture fluid. *J Microbiol Methods.* 2018;146:7-
568 12. doi: 10.1016/j.mimet.2018.01.012.
- 569 [7] Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: Current trends and future possibilities. *Br J Gen Pract.*
570 2018;68:143-144. doi: 10.3399/bjgp18X695213.
- 571 [8] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nat Med.* 2022;28:31-38. doi: 10.1038/s41591-021-01614-0.
- 572 [9] Van Veen D, Van Uden C, Blankemeier L, et al. Adapted large language models can outperform medical experts in clinical text
573 summarization. *Nat Med.* 2024. doi: 10.1038/s41591-024-02855-5.
- 574 [10] Vasey B, Ursprung S, Beddoe B, et al. Association of Clinician Diagnostic Performance With Machine Learning-Based Decision
575 Support Systems: A Systematic Review. *JAMA Netw Open.* 2021;4(3):e211276. doi: 10.1001/jamanetworkopen.2021.1276.
- 576 [11] Choi DH, Lim MH, Kim KH, et al. Development of an artificial intelligence bacteremia prediction model and evaluation of its impact
577 on physician predictions focusing on uncertainty. *Sci Rep.* 2023;13:13518. doi: 10.1038/s41598-023-40708-2.
- 578 [12] Golubnitschaja O, Kinkorova J, Costigliola V. Predictive, Preventive and Personalised Medicine as the hardcore of 'Horizon 2020':
579 EPMA position paper. *EPMA J.* 2014;5(1):6. doi: 10.1186/1878-5085-5-6.
- 580 [13] Idrees SM, Alam MA, Agarwal P, Ansari L. Effective Predictive Analytics and Modeling Based on Historical Data. In: Singh M,
581 Gupta P, Tyagi V, Flusser J, Våren T, Kashyap R, editors. *Advances in Computing and Data Sciences. ICACDS 2019.*
582 *Communications in Computer and Information Science*, vol 1046. Springer, Singapore. 2019. doi: 10.1007/978-981-13-9942-8_52.
- 583 [14] Bhavani SV, Lonjers Z, Carey KA, Afshar M, Gilbert ER, Shah NS, Huang ES, Churpek MM. The Development and Validation of a
584 Machine Learning Model to Predict Bacteremia and Fungemia in Hospitalized Patients Using Electronic Health Record Data. *Crit*
585 *Care Med.* 2020;48(11):E1020-E1028. doi: 10.1097/CCM.0000000000004556.
- 586 [15] Lee KH, Dong JJ, Jeong SJ, Chae M-H, Lee BS, Kim HJ, Ko SH, Song YG. Early detection of bacteraemia using ten clinical variables
587 with an artificial neural network approach. *J Clin Med.* 2019;8(10). doi: 10.3390/jcm8101592.
- 588 [16] Lee KH, Dong JJ, Kim S, Kim D, Hyun JH, Chae M-H, Lee BS, Song YG. Prediction of Bacteremia Based on 12-Year Medical Data
589 Using a Machine Learning Approach: Effect of Medical Data by Extraction Time. *Diagnostics.* 2022;12(1). doi:
590 10.3390/diagnostics12010102.
- 591 [17] Cheng M, Zhao X, Ding X, Gao J, Xiong S, Ren Y. Prediction of blood culture outcome using hybrid neural network model based on
592 electronic health records. *BMC Med Inform Decis Mak.* 2020;20(Suppl 3), 121. doi: 10.1186/s12911-020-1113-4.
- 593 [18] McFadden BR, Inglis TJJ, Reynolds M. Machine learning pipeline for blood culture outcome prediction using Sysmex XN-2000 blood
594 sample results in Western Australia. *BMC Infect Dis.* 2023;23(1). doi: 10.1186/s12879-023-08535-y.
- 595 [19] Lien F, Lin H-S, Wu Y-T, Chiueh T-S. Bacteremia detection from complete blood count and differential leukocyte count with
596 machine learning: complementary and competitive with C-reactive protein and procalcitonin tests. *BMC Infect Dis.* 2022;22(1). doi:
597 10.1186/s12879-022-07223-7.
- 598 [20] Mahmoud E, Dhoayan MA, Bosaeed M, al Johani S, Arabi YM. Developing machine-learning prediction algorithm for bacteremia in
599 admitted patients. *Infect Drug Resist.* 2021;14:757–765. doi: 10.2147/IDR.S293496.

- 600 [21] Garnica O, Gómez D, Ramos V, Hidalgo JI, Ruiz-Giardin JM. Diagnosing hospital bacteraemia in the framework of predictive,
601 preventive and personalised medicine using electronic health records and machine learning classifiers. *EPMA J.* 2021;12(3):365–381.
602 doi: 10.1007/s13167-021-00252-3.
- 603 [22] Murri R, De Angelis G, Antenucci L, Fiori B, Rinaldi R, Fantoni M, Damiani A, Patarnello S, Sanguinetti M, Valentini V, Posteraro
604 B, Masciocchi C. A Machine Learning Predictive Model of Bloodstream Infection in Hospitalized Patients. *Diagnostics (Basel).*
605 2024;14(4):445. doi: 10.3390/diagnostics14040445.
- 606 [23] Ratzinger F, Haslacher H, Perkmann T, Pinzan M, Anner P, Makristathis A, Burgmann H, Heinze G, Dorffner G. Machine learning
607 for fast identification of bacteraemia in SIRS patients treated on standard care wards: a cohort study. *Sci Rep.* 2018;8(1). doi:
608 10.1038/s41598-018-30236-9.
- 609 [24] Rahmani K, Garikipati A, Barnes G, Hoffman J, Calvert J, Mao Q, Das R. Early prediction of central line associated bloodstream
610 infection using machine learning. *Am J Infect Control.* 2022;50(4):440–445. doi: 10.1016/j.ajic.2021.08.017.
- 611 [25] Beeler C, Dbeibo L, Kelley K, Thatcher L, Webb D, Bah A, Monahan P, Fowler NR, Nicol S, Judy-Malcolm A, Azar J. Assessing
612 patient risk of central line-associated bacteremia via machine learning. *Am J Infect Control.* 2018;46(9):986–991. doi:
613 10.1016/j.ajic.2018.02.021.
- 614 [26] Zhou T, Ren Z, Ma Y, He L, Liu J, Tang J, Zhang H. Early identification of bloodstream infection in hemodialysis patients by
615 machine learning. *Heliyon.* 2023;9(7). doi: 10.1016/j.heliyon.2023.e18263.
- 616 [27] Yoo J, Kim SH, Hur S, Ha J, Huh K, Cha WC. Candidemia Risk Prediction (CanDETEC) Model for Patients With Malignancy:
617 Model Development and Validation in a Single-Center Retrospective Study. *JMIR Med Inform.* 2021;9(7), e24651. doi:
618 10.2196/24651.
- 619 [28] Mooney C, Eogan M, Ní Áinle F, Cleary B, Gallagher JJ, O’Loughlin J, Drew RJ. Predicting bacteraemia in maternity patients using
620 full blood count parameters: A supervised machine learning algorithm approach. *Int J Lab Hematol.* 2021;43(4):609–615. doi:
621 10.1111/ijlh.13434.
- 622 [29] Su M, Chen H, Qiu J, Huang J. Four Biomarkers-Based Artificial Neural Network Model for Accurate Early Prediction of Bacteremia
623 with Low-level Procalcitonin. *Ann Clin Lab Sci.* 2021;51(3):408–414.
- 624 [30] Wu L, Xia D, Xu K. Multi-Clinical Factors Combined with an Artificial Intelligence Algorithm Diagnosis Model for HIV-Infected
625 People with Bloodstream Infection. *Infect Drug Resist.* 2023;16:6085–6097. doi: 10.2147/IDR.S423709.
- 626 [31] Choi DH, Hong KJ, Park JH, Shin SD, Ro YS, Song KJ, Kim KH, Kim S. Prediction of bacteremia at the emergency department
627 during triage and disposition stages using machine learning models. *Am J Emerg Med.* 2022;53:86–93. doi:
628 10.1016/j.ajem.2021.12.065.
- 629 [32] Choi DH, Lim MH, Kim KH, Shin SD, Hong KJ, Kim S. Development of an artificial intelligence bacteremia prediction model and
630 evaluation of its impact on physician predictions focusing on uncertainty. *Sci Rep.* 2023;13(1). doi: 10.1038/s41598-023-40708-2.
- 631 [33] Boerman AW, Schinkel M, Meijerink L, van den Ende ES, Pladet LC, Scholtemeijer MG, Zeeuw J, van der Zaag AY, Minderhoud
632 TC, Elbers PWG, Wiersinga WJ, de Jonge R, Kramer MH, Nanayakkara PWB. Using machine learning to predict blood culture
633 outcomes in the emergency department: a single-centre, retrospective, observational study. *BMJ Open.* 2022;12(1):e053332. doi:
634 10.1136/bmjopen-2021-053332.
- 635 [34] Chang Y-H, Hsiao C-T, Chang Y-C, Lai H-Y, Lin H-H, Chen C-C, Hsu L-C, Wu S-Y, Shih H-M, Hsueh P-R, Cho D-Y. Machine
636 learning of cell population data, complete blood count, and differential count parameters for early prediction of bacteremia among

- 637 adult patients with suspected bacterial infections and blood culture sampling in emergency departments. *J Microbiol Immunol Infect.*
638 2023;56(4):782–792. doi: 10.1016/j.jmii.2023.05.001.
- 639 [35] Schinkel M, Boerman AW, Bennis FC, Minderhoud TC, Lie M, Peters-Sengers H, Holleman F, Schade RP, de Jonge R, Wiersinga
640 WJ, Nanayakkara PWB. Diagnostic stewardship for blood cultures in the emergency department: A multicenter validation and
641 prospective evaluation of a machine learning prediction tool. *EBioMedicine.* 2022;82. doi: 10.1016/j.ebiom.2022.104176.
- 642 [36] Goh V, Chou Y-J, Lee C-C, Ma M-C, Wang WYC, Lin C-H, Hsieh C-C. Predicting Bacteremia among Septic Patients Based on ED
643 Information by Machine Learning Methods: A Comparative Study. *Diagnostics.* 2022;12(10). doi: 10.3390/diagnostics12102498.
- 644 [37] Tsai W-C, Liu C-F, Ma Y-S, Chen C-J, Lin H-J, Hsu C-C, Chow JC, Chien Y-W, Huang C-C. Real-time artificial intelligence system
645 for bacteremia prediction in adult febrile emergency department patients. *Int J Med Inform.* 2023. doi:
646 10.1016/j.ijmedinf.2023.105176.
- 647 [38] Roimi M, Neuberger A, Shrot A, Paul M, Geffen Y, Bar-Lavie Y. Early diagnosis of bloodstream infections in the intensive care unit
648 using machine-learning algorithms. *Intensive Care Med.* 2020;46(3):454–462. doi: 10.1007/s00134-019-05876-8.
- 649 [39] Van Steenkiste T, Ruyssinck J, De Baets L, Decruyenaere J, De Turck F, Ongenae F, Dhaene T. Accurate prediction of blood culture
650 outcome in the intensive care unit using long short-term memory neural networks. *Artif Intell Med.* 2019;97:38–43. doi:
651 10.1016/j.artmed.2018.10.008.
- 652 [40] Boner Z, Moore CC. Deep Learning Risk Prediction of Bloodstream Infection in the Intensive Care Unit. In: *KDD Undergraduate*
653 *Consortium*, August 14-18, 2022, Washington, D.C., USA: ACM; 2022.
- 654 [41] Pai K-C, Wang M-S, Chen Y-F, Tseng C-H, Liu P-Y, Chen L-C, Sheu R-K, Wu C-L. An artificial intelligence approach to
655 bloodstream infections prediction. *J Clin Med.* 2021;10(13). doi: 10.3390/jcm10132901.
- 656 [42] Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using
657 supervised machine learning. *J Crit Care.* 2018;45:156-162. doi: 10.1016/j.jcrc.2018.02.010.
- 658 [43] Yuan S, Sun Y, Xiao X, Long Y, He H. Using Machine Learning Algorithms to Predict Candidaemia in ICU Patients With New-Onset
659 Systemic Inflammatory Response Syndrome. *Front Med.* 2021;8. doi: 10.3389/fmed.2021.720926.
- 660 [44] Mozes B, Milatiner D, Block C, Blumstein Z, Halkin H. Inconsistency of a model aimed at predicting bacteremia in hospitalized
661 patients. *J Clin Epidemiol.* 1993;46(9):1035-1040. doi: 10.1016/0895-4356(93)90171-v.
- 662 [45] Dargère S, Cormier H, Verdon R. Contaminants in blood cultures: importance, implications, interpretation and prevention. *Clin*
663 *Microbiol Infect.* 2018;24(9):964-969. doi: 10.1016/j.cmi.2018.03.030.
- 664 [46] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis
665 or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594. doi: 10.1136/bmj.g7594.
- 666 [47] Montgomery DC, Runger GC. *Applied Statistics and Probability for Engineers.* 5th ed. Hoboken, NJ: John Wiley & Sons; 2013.
- 667 [48] Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. *Am Stat.* 1988;42(1):59-66.
- 668 [49] Coburn B, Morris AM, Tomlinson G, Detsky AS. Does This Adult Patient With Suspected Bacteremia Require Blood Cultures?
669 *JAMA.* 2012;308(5):502–511. doi: 10.1001/jama.2012.8262.
- 670 [50] Eliakim-Raz N, Bates DW, Leibovici L. Predicting bacteraemia in validated models--a systematic review. *Clin Microbiol Infect.*
671 2015;21(4):295–301. doi: 10.1016/j.cmi.2015.01.023.

- 672 [51] Laukemann S, Kasper N, Kulkarni P, Steiner D, Rast AC, Kutz A, Felder S, Haubitz S, Faessler L, Huber A, Fux CA, Mueller B,
673 Schuetz P. Can We Reduce Negative Blood Cultures With Clinical Scores and Blood Markers? Results From an Observational Cohort
674 Study. *Medicine*. 2015;94(49):e2264. doi: 10.1097/MD.0000000000002264.
- 675 [52] Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? arXiv preprint
676 arXiv:2207.08815. 2022.
- 677 [53] Bates DW, Goldman L, Lee TH. Contaminant blood cultures and resource utilization. The true consequences of false-positive results.
678 *JAMA*. 1991;265(3):365–369.
- 679 [54] Zwang O, Albert RK. Analysis of strategies to improve cost effectiveness of blood cultures. *J Hosp Med*. 2006;1(5):272–276. doi:
680 10.1002/jhm.115.
- 681 [55] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: *Proceedings of the 30th International Conference*
682 *on Neural Information Processing Systems*; 2017.
- 683 [56] Bopche R, Gustad LT, Afset JE, Damås JK, Nytrø Ø. Predicting In-Hospital Death from Derived EHR Trajectory Features. *Stud*
684 *Health Technol Inform*. 2024;310:269–273. doi: 10.3233/SHTI230969.
- 685 [57] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I. Explainable ai for trees:
686 From local explanations to global understanding. arXiv preprint arXiv:1905.04610. 2019.
- 687 [58] Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*.
688 2020;2:56–67. doi: 10.1038/s42256-019-0138-9.
- 689
690
691
692
693
694
695
696

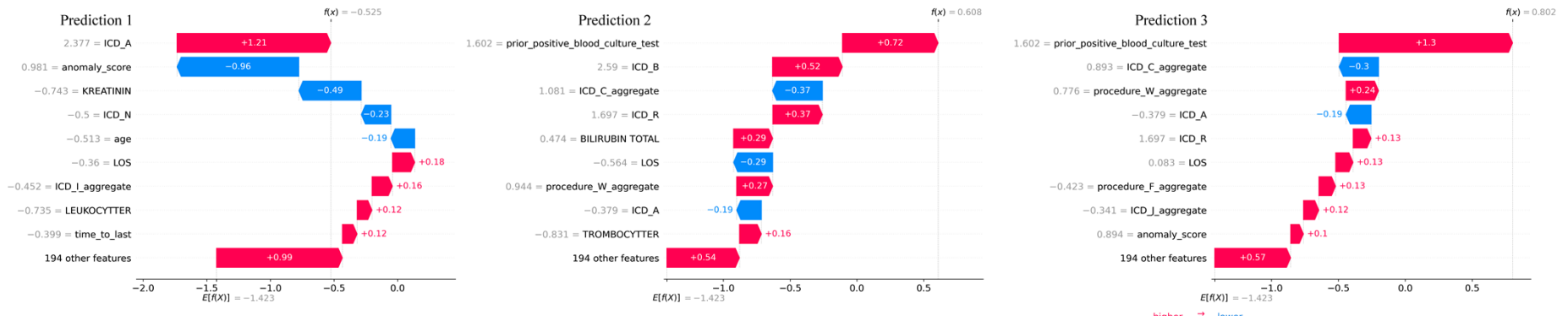
Table 1. The key characteristics of the studies using ML for BSI prediction

Setting	First author, year	Target condition	No. Patients	Data source	Prevalence	ML models	Key predictors
Inpatients	Bhavani et al. (2020) [14]	Bacteremia and Fungemia	76688	EHRs, University of Chicago Hospital (2008-2018) and Loyola University Medical Center, USA (2007-2017)	Bacteremia 7.7%, Fungemia 0.7%	LR, GBM	Time from admission to BC, Temperature, Age, HR, Prior Bacteremia/Fungemia, WBC, BUN, Glucose, DBP, SBP, PPI, RR
	Lee et al. (2019) [15]	Bacteremia	13402	EHRs, Gangnam Severance Hospital, Seoul, Republic of Korea, (2008-2012)	7.9%	MLP, SVM, RF	ALP, PLT, Temperature, SBP, WBC, ICU stay, CRP, CVC, Age, PT, Hospital days to BC, HR, Gender, Antibiotics, RR, Creatinine
	Lee et al. (2022) [16]	Bacteremia	622771 samples	EHRs, Sinchon and Gangnam Severance Hospitals, Republic of Korea, (2007-2018)	6.2%	MLP, RF, XGB	PLT, Monocyte, Neutrophil, Bilirubin, Albumin, and Hospital stay, BUN, ALP, RR, PR, DBP, TP, WBC, PT, Hb, CRP, Creatinine, ALT, AST, Sodium, Chloride, ESR
	Cheng et al. (2020) [17]	Bacteremia	28043	EHRs, Zhengzhou University Hospital, China, (2017-2018)	10%	LR, NB, SVM, ADT, CNN, BiLSTM, ABiLSTM+ DAE	Textual chief complaints, Admission records, and Laboratory biochemical indicators.
	McFadden et al. (2023) [18]	Bacteremia	10965 samples	CBC/DC, CPD, Sir Charles Gairdner Hospital, Western Australia (2018-2020)	7.58%	RF, XGB	CBC, DIFF, and CPD
	Lien et al. (2022) [19]	Bacteremia	366586 samples	EHRs, CBC/DC, Linkou Chang Gung Memorial Hospital (CGMH) in Taiwan, (2014-2019)	8.2%	RF, LR	CBC/DC, CRP, and PCT

	Mahmoud et al. (2021) [20]	Bacteremia	7157	EHRs, King Abdulaziz Medical City, Riyadh, Saudi Arabia (2017-2019)	11.4%	NN, RF, LR, DT, NB, SVM	Age, Antibiotics use, Surgery within 14 days, CVC, length of hospitalization before BC, RR, SBP, Temperature, DBP, HR, WBC, Sodium, PLT, Albumin, Creatinine, Lactic acid level.
	Garnica et al. (2021) [21]	Bacteremia	4357	EHRs, Microbiological data, Hospital Universitario de Fuenlabrada, Madrid, Spain, (2005-2015)	51.3%	SVM, RF, KNN	The number of days in ICU before BC extraction, presence of Catheters, Chronic Respiratory disease, Fever, Age, CRP, PLT.
	Murri et al. (2024) [22]	HA-BSI samples	5660	Generator Center at the Fondazione Policlinico Universitario A. Gemelli IRCCS (FPG), Rome, Italy (2016-2019)	33.6%	LR	Time BSI > 12 days, Procalcitonin > 1 ng/mL, Presence of a CVC, PLT, Hypotension, BUN, Presence of urinary catheter, Fever, Tachycardia, Altered mental status, Age, Bilirubin, Creatinine
Inpatients with SIRS	Ratzinger et al. (2018) [23]	Bacteremia	466	EHRs Vienna General Hospital, Austria, (2011-2012)	28.8%	RF, ANN, ENR	PCT, LBP, Albumin, Bilirubin
Inpatients with CVC	Rahmani et al. (2022) [24]	CLABSIs	27619	EHRs, a proprietary national longitudinal EHR repository, Houston, Texas, USA (2015-2020)	1%	XGB, DT, LR	Temperature, HGB, comorbidities, Age, WBC, Race, Neutrophil.
	Beeler et al. (2018) [25]	CLABSIs	70218	EHRs, Indiana University Health Academic Health Center, USA, (2013-2016)	0.6%	RF, LR	Age, Gender, history of CLABSI, CHG (Chlorhexidine Gluconate) Bathing Non-compliant Days, Line days.
HD patients	Zhou et al. (2023) [26]	Bacteremia	391	EHRs, Department of Nephrology, Affiliated Hospital of North Sichuan Medical College, Sichuan Province, China, (2018-2022)	18.9%	LR, SVM, DT, RF, XGB	PCT, Temperature, Non-arteriovenous fistula dialysis access, NLR, Leukocyte, dialysis duration, LMR, Albumin, Neutrophil, PLT, Age, DBP, CRP, PLR, ALP, SBP, HR, history of BSI,
Cancer patients	Yoo et al. (2021) [27]	Candidemia	34574	EHRs, academic single hospital in Seoul, Republic of Korea, (2010-2018)	0.6%	LR, ANN, RF, GBM, AML	Variables reflecting the dynamic status of patients with cancer, including blood urea nitrogen level, 7-day variance of RR, Total bilirubin level, 7-day variance of SBP, Body weight.

Maternity patients	Mooney et al. (2020) [28]	Bacteremia	129	CBC parameters, Rotunda Hospital, Ireland (2019)	3%	CART, LDA, KNN, SVM, RF	NLR, CBC parameters.
Patients with PCT ≤ 2.0 ng/ml	Su et al. (2021) [29]	Bacteremia	931	EHRs, Mindong Hospital Affiliated to Fujian Medical University, China, (2014-2020)	47%	ANN, KNN, LR, RF, SVM, and NB.	Interleukin-6, PCT, D-dimer, Lactic acid, Leukocytes, Neutrophil, and PLT.
HIV patients	Wu et al. (2023) [30]	Bacteremia	498	EHRs, Wenzhou Central Hospital, China, (2014-2021)	34.3%	SVM, ANN, GBM, GLM, MDA, PLR, NB, RF	Low Hb, CD4+T cell, PLT, LDH, BUN, splenomegaly, absence of ART treatment, Strip shadow, Nodular shadow, and Shock.
ED	Choi et al. (2022) [31]	Bacteremia	24768	EHRs, An urban tertiary referral hospital, Republic of Korea, (2016-2018)	12%	XGB, RF, LR	Chief complaint, Age, Temperature, HR, and DBP at triage stage. Neutrophils, PLT, CRP, Chief complaints, and Creatinine at disposition stage.
	Choi et al. (2023) [32]	Bacteremia	15362	EHRs, Seoul National University Hospital, Seoul National University Bundang Hospital, Republic of Korea, (2016-2018)	10.9%	BNN	Age, HR, Temperature, DBP, History of chills, Ambulance use
	Boerman et al. (2022) [33]	Bacteremia	4885	EHRs, Amsterdam UMC, location VU University Medical Center, NL, (2018-2020)	12.2%	GBT, LR	Bilirubin, Urea, lymphocyte, Pulse rate, CRP, Neutrophil, age, Temperature, DBP, Potassium, Glucose, Thrombocytes, Creatinine, ALP, SBP, Organ damage
	Chang et al. (2023) [34]	Bacteremia	20636	EHRs, CPD, CBC/DC, China Medical University Hospital, Taiwan, (2021-2022)	10.4%	CatBoost, LGBM, XGB, RF, LR	Demographics, CPD, CBC/DC
	Schinkel et al. (2022) [35]	Bacteremia	6421	EHRs, Amsterdam UMC, (VUMC, AMC, ZMC, and BIDMC), NL, (2016-2021)	5.4% - 12.3%	XGB, LR	Temperature, Creatinine, CRP, Lymphocytes, DBP, Bilirubin, Thrombocytes, Neutrophils, ALP, HR, SBP, Leukocytes, Glucose, Age, Potassium, BUN, Sodium, monocytes
ED patients with SIRS	Goh et al. (2022) [36]	Bacteremia	40395	EHRs, National Cheng Kung University Hospital, Taiwan, (2015-2019)	10%	LR, SVM, RF	Age, Gender, COPD, Uncomplicated DM, Hemato-oncology, WBC, Band cell, Platelet, Temperature, HR, mild liver disease, Mean arterial pressure, RR, GSC

ED patients with fever	Tsai et al. (2023) [37]	Bacteremia	3669	EHRs, Chi Mei Medical Center, Taiwan, (2017-2020)	13.8%	RF, LR, MLP, XGB, LGBM	Hypertension, Gender, Temperature, DM, Age, CRP, PLT, WBC, Malignancy, Eosinophil, HR, BMI, Hb, RR, SBP, DBP, Band, CKD, Liver Cirrhosis, COPD, GCS
ICU patients	Roimi et al. (2020) [38]	Bacteremia	3372	EHRs, BIDMC, Boston, Massachusetts, USA, (2008-2012), ICU of Rambam Healthcare Campus (RHCC), Israel, (2013-2017)	ICU acquired: 6.4% (BIDMC), 15.9% (RHCC)	RF, XGB	Time duration (days) between sampling time and last defecation, Time duration (hours) between sampling time and the maximum BUN (mg/dL) value measured during the 5 days prior to sampling, Length of stay (days) between sampling time and ICU admission, The minimal weight (kg) during the 5 days prior to sampling, The time duration between sampling time and the maximum MCHC (g/dL) during the 5 days prior to sampling
	Van Steenkiste et al. (2019) [39]	Bacteremia	2177	EHRs, ICU, Ghent University Hospital, Belgium, (2013-2015)	10.5%	BiLSTM, ANN, SVM, KNN, LR	Temperature, Thrombocytes, Leukocytes, CRP, sepsis-related organ failure assessment, HR, RR, PT, and mean systemic arterial pressure.
	Boner et al. (2022) [40]	Bacteremia	6557	EHRs, ICU, University of Virginia, USA, (2011-2015)	13.3%	FNN, GRU, CNN, LR	Temperature, BUN, BP, HR, Albumin, PLT, Chloride, Creatinine, Chloride, and Phosphorus.
	Pai et al. (2021) [41]	Bacteremia	4275	EHRs, Taichung Veterans General Hospital ICU, Taiwan, (2015-2019)	13.8%	LR, SVM, MLP, RF, XGB	ALP, CVC period, prothrombin time, PLT, Albumin, Apache II score, Age, foley
ICU patients with CVC	Parreco et al. (2018) [42]	CLABSIs	57786 admissions	MIMIC-III database, USA, (2001-2012)	1.5%	LR, GBT, DL	Severity of illness scores (like SAPS II, APS III, and OASIS) and comorbidities.
ICU patients with new-onset SIRS	Yuan et al. (2021) [43]	Candidemia	7932	EHRs, Peking Union Medical College Hospital, The Affiliated Hospital of Qingdao University, The First Affiliated Hospital of Fujian Medical University, China, (2013-2017)	1%	XGB, SVM, RF, ET, LR	Colonization, Diabetes, AKI, total number of parenteral nutrition days, history of fungal infection, CRRT days, Abdominal surgery, BDG, days of mechanical ventilation, Length of hospital and ICU stay, days of CVC



Feature Descriptions:

- ICD_A - Counts of ICD-10 Chapter I (A00-A99) Infectious and parasitic diseases, in the recent episode
- ICD_B - Counts of ICD-10 Chapter I (B00-B99) Infectious and parasitic diseases, in the recent episode
- BILIRUBIN TOTAL - Total Bilirubin
- KREATININ - Creatinine
- TROMBOCYTTER - Thrombocytes
- ICD_N - Counts of ICD-10 codes Chapter XIV (N00-N99) Genitourinary diseases, in the recent episode
- LOS - Length of stay at hospital till BC test
- ICD_I_aggregate - Counts of ICD-10 codes Chapter IX (I00-I99) Diseases of circulatory system, in the history
- LEUKOCYTTER - Leukocytes
- time_to_last - Time to the most recent admission
- prior_positive_blood_culture_test - Counts of prior positive results for BC tests
- procedure_W_aggregate - Counts of procedures on female reproductive organs in the history
- procedure_F_aggregate - Counts of procedures on cardiovascular systems in the history
- ICD_C_aggregate - Counts of ICD-10 Chapter II (C00-C99) Cancer related codes, in the history
- ICD_R - Counts of ICD-10 Chapter XVIII (R00-R99) Symptoms, Signs, not elsewhere classified in the recent episode
- prior_positive_hal - Counts of positive microbiology test results for throat secret
- ICD_J_aggregate - Counts of ICD-10 Chapter X (J00-J99) respiratory diseases, in the history
- anomaly_score - Scores of Isolation Forest model

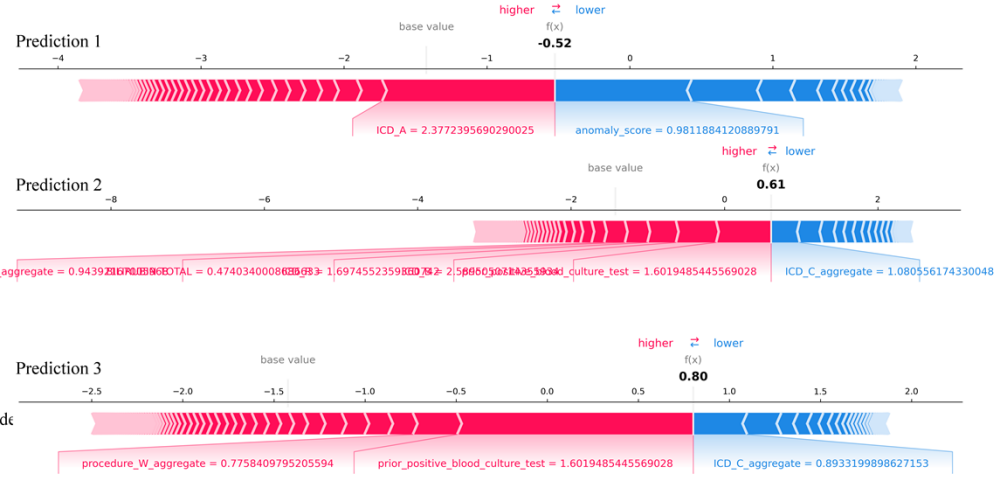


Figure 4. Waterfall and Force plots of first three predictions from the testing set.

Figure 4. Waterfall and Force plots for the first three predictions from the test set.

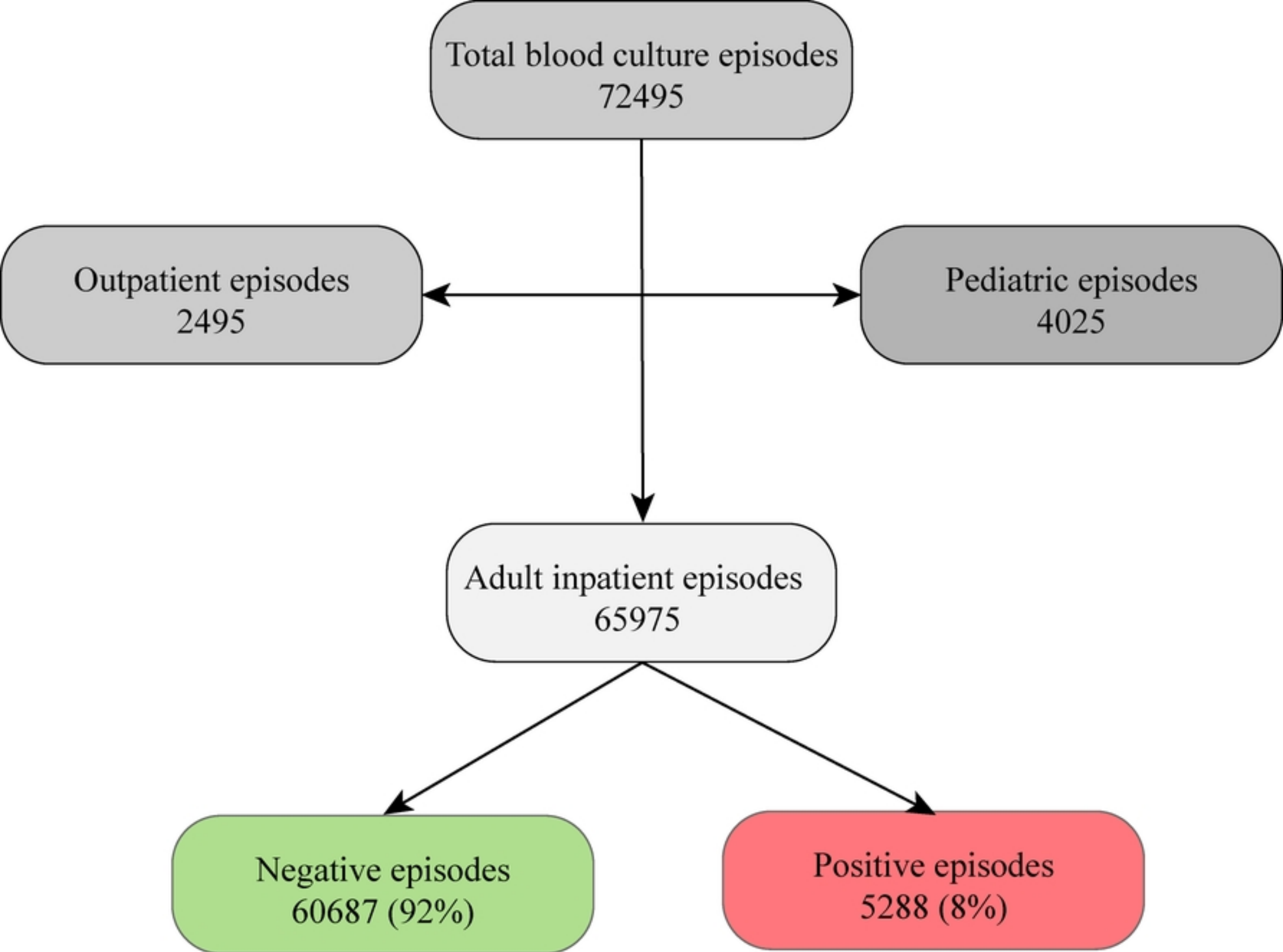


Figure 2

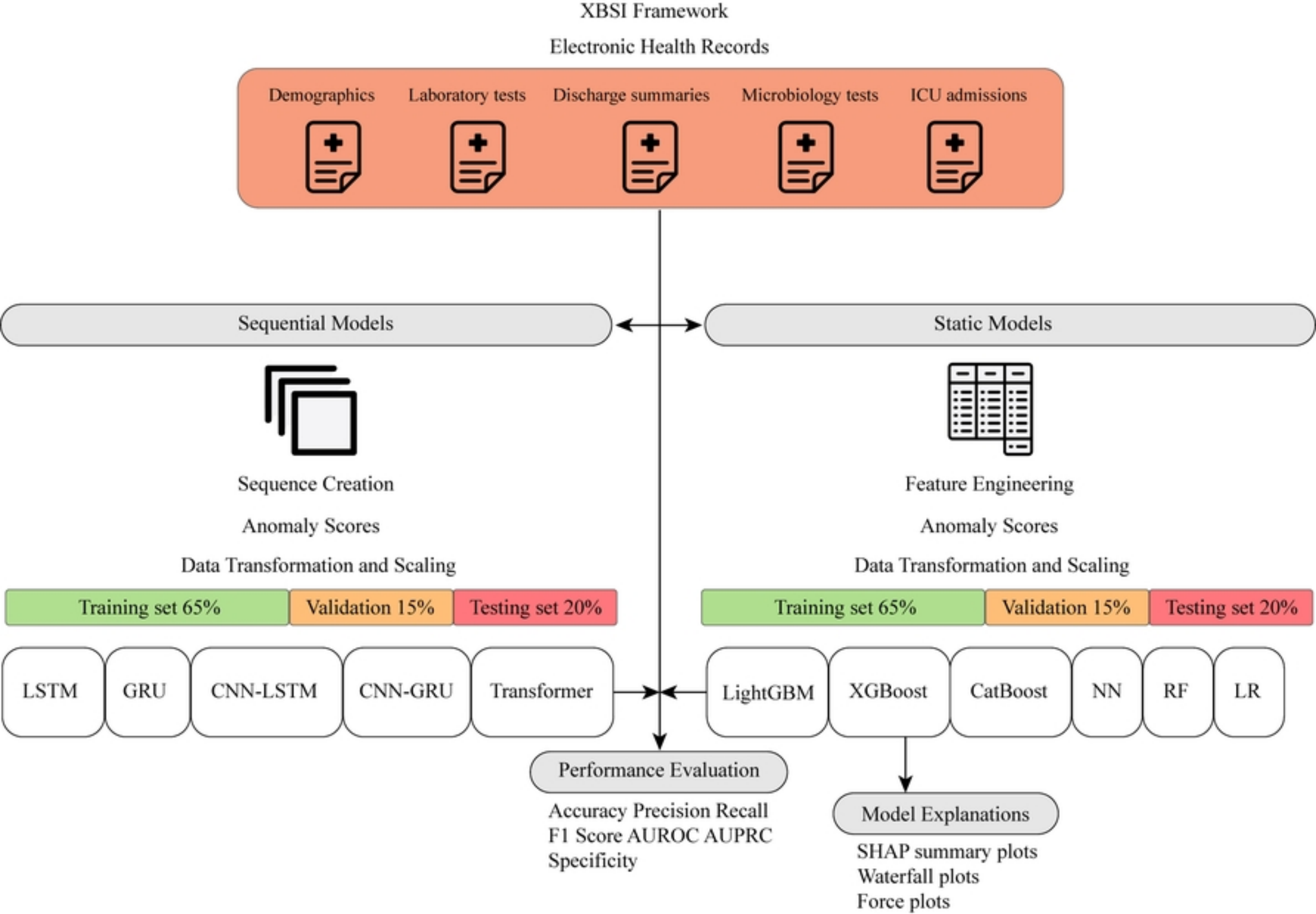


Figure 1

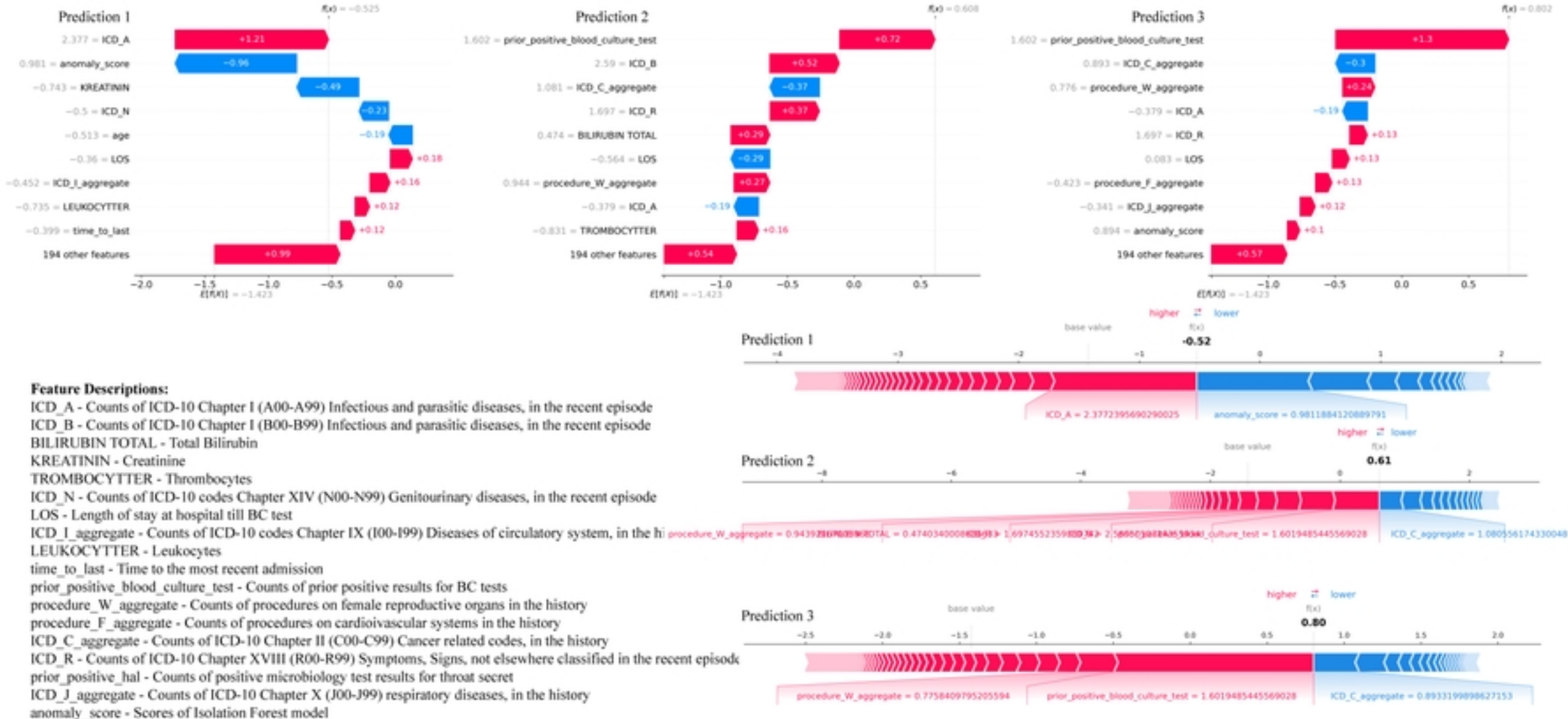


Figure 4. Waterfall and Force plots of first three predictions from the testing set.

Figure 3. SHAP Summary Plots for XGBoost Model Feature Importance.

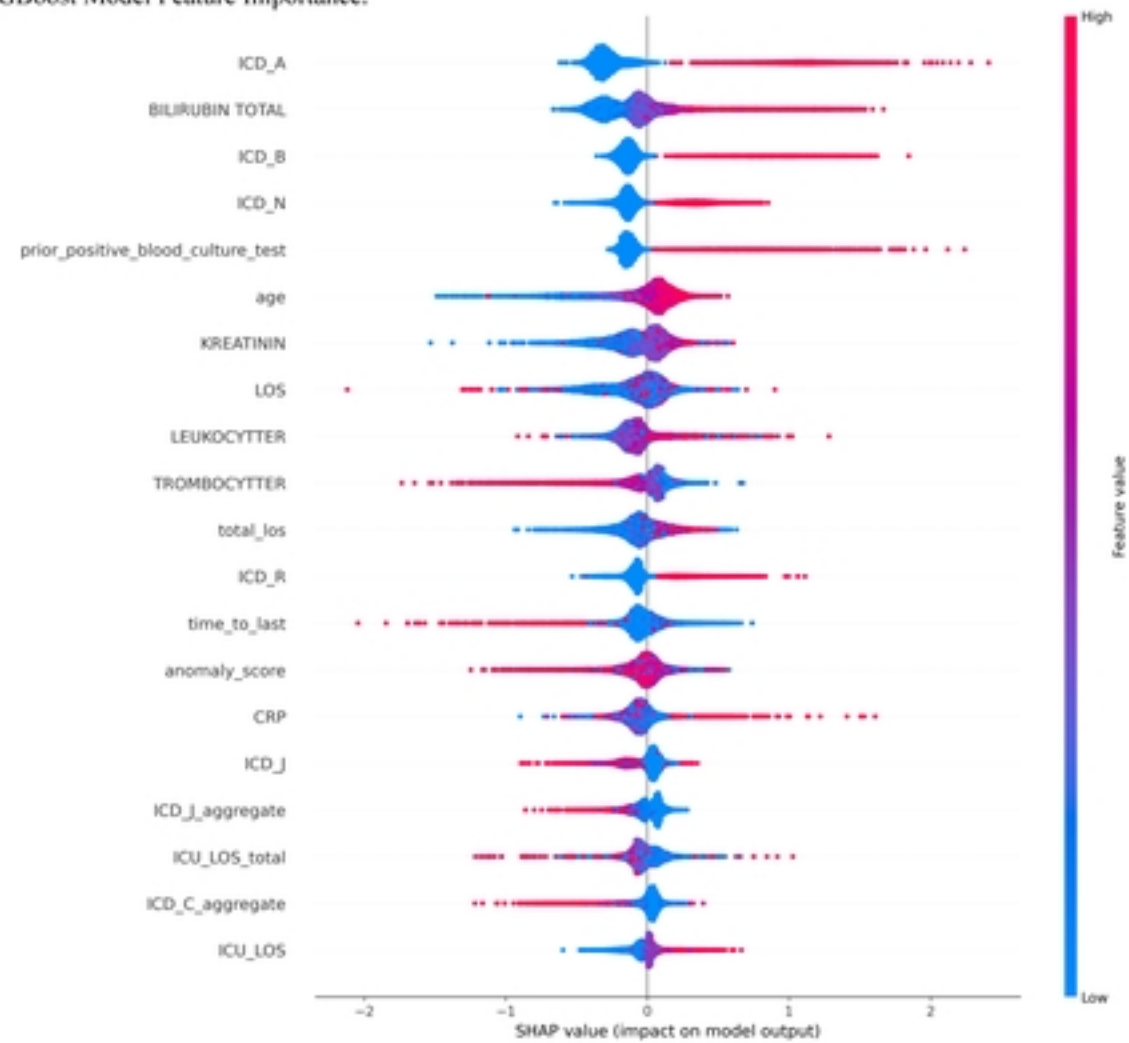
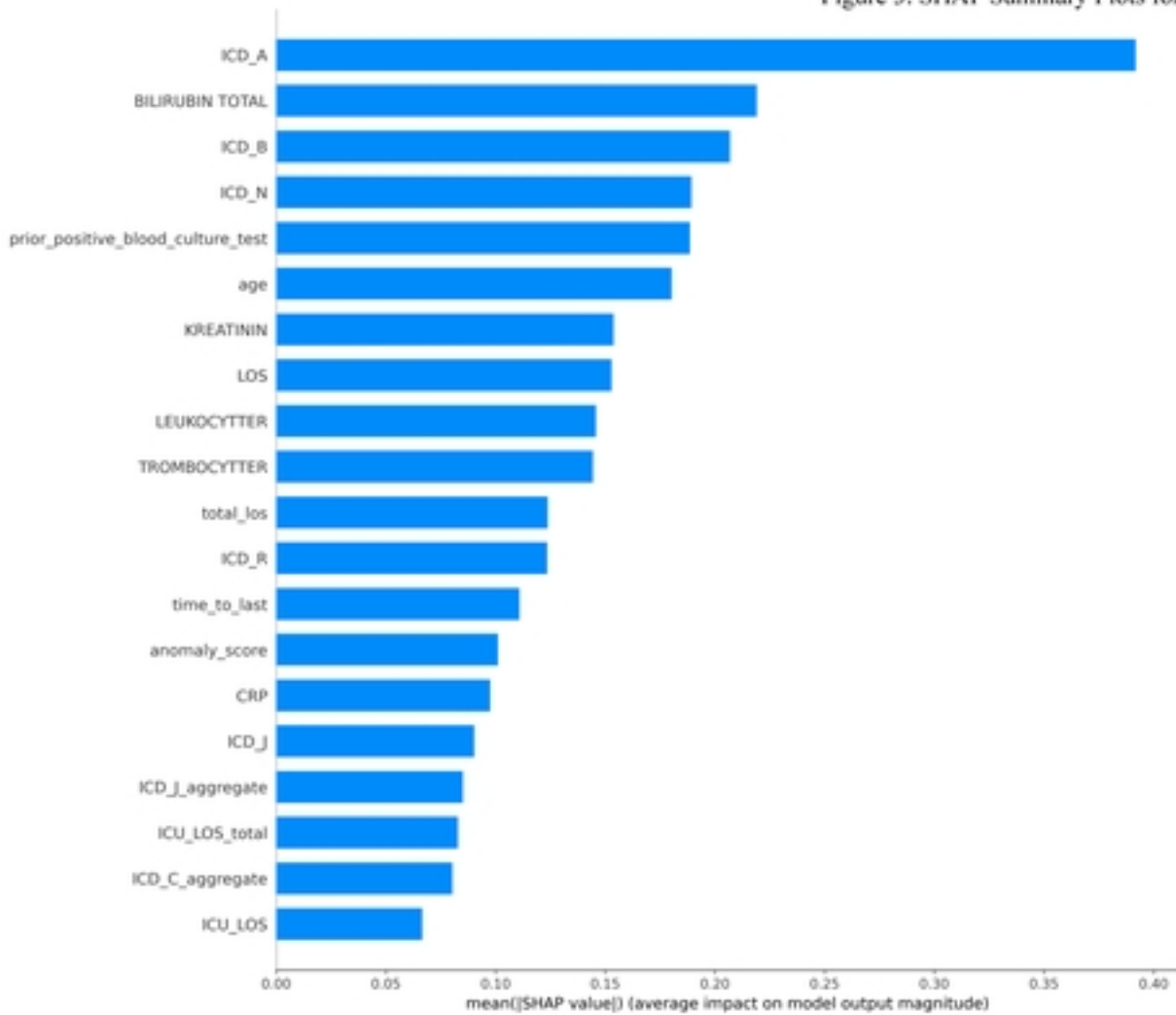


Figure 3