A Deep Learning Approach for Culture-Free Bacterial Meningitis Diagnosis and ICU Outcome Prediction

Rui Chen[1†], Yu Cai[1†], Sitong Zhang[1], Zirong Huo[1], Mingming Song[1], Wenqing Li[1], Dongyan Yang[1], Xi Zhang[1,2]*

1 School of Life Science, Northwest University, Xi'an, Shaanxi, 710069, China.

2 School of Professional Studies, Northwestern University, Evanston, IL 60201, US.

† Rui Chen and Yu Cai have contributed equally to this work.

* Corresponding author

Rui Chen: yourangle@126.com

Yu Cai: caiyu1316@163.com

Sitong Zhang: zstongoo@163.com

Zirong Huo: hzr1661864339@163.com

Mingming Song: 1245602472@qq.com

Wenqing Li: 927395328@qq.com

Dongyan Yang: 202322544@stumail.nwu.edu.cn

Xi Zhang: xzhang19@nwu.edu.cn

Abstract

Background: Cerebrospinal fluid (CSF) culture is essential for diagnosing neuroinfectious diseases such as meningitis, where clinical symptoms alone cannot reliably confirm the diagnosis or differentiate bacterial from non-bacterial causes. However, traditional CSF cultures are time-consuming and often lack sensitivity, particularly after antibiotic treatment. This study applies machine learning to expedite diagnosis and enhance the accuracy of outcome predictions for patients with suspected neuroinfectious diseases.

Methods: Training and validation datasets were derived from the MIMIC-III and MIMIC-IV databases, respectively, focusing on patients undergoing CSF bacterial culture. CNN, FCN, and RF classifiers were trained using integrated test results and clinical data to predict CSF culture outcomes. CSF culture result labels were redefined based on clinical guidelines for bacterial meningitis diagnosis to address false negatives caused by prior antibiotic treatment. Unstructured text data were encoded using BioBERT embeddings, combined with structured laboratory data, and processed through CNN, FCN, or LSTM layers to develop multimodal neural networks for predicting clinical outcomes. Class imbalance was mitigated, and text data sources were optimized for interchangeability.

Results: For CSF culture diagnostic predictions, 9261 cases were analyzed, including 245 initial culture-positive and 9016 culture-negative results. Label correction increased the positive cases to 404, with 8857 negatives. The FCN model showed the best performance post-label correction, achieving an AUC of 0.891, accuracy of 0.867, sensitivity of 0.833, and specificity of 0.868. In survival prediction analysis (5795 cases: 5500 survivors, 295 in-hospital deaths), the optimized multimodal neural network with CNN layers achieved an AUC of 0.933, accuracy of 0.850, sensitivity of 0.847, and specificity of 0.850. Models

trained without text data showed diminished performance. Independent validation datasets confirmed the model's robustness, yielding comparable results.

Conclusions: The multimodal machine learning models demonstrated strong efficacy in predicting CSF culture outcomes, supporting the diagnosis of bacterial meningitis, and in forecasting ICU survival outcomes. Integrating these models into a clinical decision support system could empower ICU clinicians with timely, data-driven insights, enabling culture-independent diagnostics and enhancing patient management.

Keywords: Cerebrospinal Fluid; Bacterial Meningitis; Multimodal neural networks; ICU Survival; BioBERT

Background

Neuroinfectious diseases, which include infections of the central nervous system such as bacterial and viral meningitis and viral encephalitis, can lead to significant complications if not diagnosed and treated promptly. Timely and accurate diagnosis is essential to guide the most effective clinical management [1]. Comparing to other types of neuroinfectious disease, bacterial meningitis (BM) typically has a rapid onset and high mortality rate which is largely attributed to delayed diagnosis and treatment [2]. Improving diagnostic accuracy through more refined laboratory analyses and advanced techniques like machine learning models would directly enhance the effectiveness of treatment for BM patients [3].

When acute BM is suspected, blood samples should be analyzed and an immediate lumbar puncture is recommended to assess whether the CSF composition aligns with the clinical diagnosis [4]. The gold standard diagnosis of BM primarily relies on a combination of clinical findings and CSF analysis, with CSF culture serving as the definitive confirmation [5]. It is notable that CSF culture has time-consuming limitations and a high risk of false negatives [6, 7]. Because of these limitations, it is encouraged that CSF culture result should be processed alongside other lab test results and clinical evidences to improve diagnostic accuracy [8]. However, it is always challenging for clinicians to quickly interpret analyzing numerous subtle clinical indicators, lab test results, and patient-specific factors, as they must consider multiple thresholds and pathogen-specific exceptions. For example, the CSF leukocyte count is a key factor in distinguishing BM from non-BM, but interpreting it requires careful analysis [5]. A count of ≥1000 cells/μL, particularly with a predominance of neutrophils, strongly suggests BM. But if the count is below 1000 cells/μL with more lymphocytes, it is more likely viral meningitis or tuberculous meningitis. An exception is BM caused by L. monocytogenes, where about 60% of cases show fewer than 1000 cells/μL, often with lymphocytic predominance.

In addition to diagnosing BM, assessing a patient's survival offers more direct and informative insights into the potential outcomes, which also require greater clinical precision and attention from clinicians [9]. A survival prediction model utilizes comprehensive patient data to develop a detailed understanding of the patient's current condition and anticipated future trajectory, empowering clinicians to make more informed and timely decisions regarding interventions and care strategies. Unstructured data, particularly text data from clinical notes, can be a valuable input. Unlike structured data (such as lab results), unstructured data (such as the chief complaint) often contains rich, nuanced details about a patient's condition and prior care [10]. This is particularly valuable for ICU patients, who are often critically ill and may be unable to provide direct input themselves. Natural Language Processing (NLP) tools, especially advanced encoders like BERT (Bidirectional Encoder Representations from Transformers) family (such as BioBERT) [11, 12] and models based on Recurrent Neural Network and Long Short-Term Memory (LSTM) [13, 14], have proven effective in capturing implicit relationships between terms and phrases related to patient prognosis by processing unstructured text data. To date, there has been very little research leveraging NLP encoders in combination with other data modalities for predicting clinical outcomes such as survival, especially in critical care settings. This gap presents a unique opportunity to explore its potential in enhancing predictive accuracy for patient survival analysis in ICU environments.

In this study, the primary objective was to develop and evaluate robust machine learning models capable of identifying critical patterns within complex laboratory test data and clinical evidence that may not be immediately evident to less experienced clinicians. This approach is particularly valuable when diagnostic information is comprehensive but challenging to interpret, especially for predicting the likelihood of positive CSF cultures. By integrating and transforming unstructured text data enriched with clinical notes by BioBERT

encoder, the model demonstrated its potential to assist clinicians in predicting clinical outcomes. The tools developed in this study are expected to provide culture-independent, data-driven insights, expediting the diagnostic process and supporting early intervention strategies.

Methods

Data Source

This study utilized data from the MIMIC-IV (v2.2) database, a publicly available repository containing de-identified health data from critical care patients between 2008 and 2019 [15, 16]. The dataset includes structured (e.g., lab results, vital signs) and unstructured data (e.g., clinical notes, discharge summaries) for 299712 patients across 431231 admissions and 73181 ICU stays. MIMIC-III (v1.4), which contains data from 2001 to 2012, was used for independent validation [17]. To avoid overlap, a subset of MIMIC-III (v1.4-subset) covering 2001–2008 was created and served as an independent validation cohort [18]. Database access required completion of the NIH “Protecting Human Research Participants” course (Certification ID: 10093146 and 12810933). This analysis adheres to the Reporting of Studies Conducted Using Observational Routinely Collected Health Data guidelines for administrative claims data [19].

Textual Data Processing

Unstructured text features (chief complaint, ICD text, past medical history, history of present illness, allergies, and family history) were transformed into numerical vectors using BioBERT, a biomedical adaptation of BERT pre-trained on PubMed articles (Fig. S1). Text tokenization converted inputs into semantic embeddings for integration with structured data. For MIMIC-III, where chief complaint data was absent, the biomedical-ner-all tool was used to extract equivalent information from clinical notes (Fig. S2).

Data Extraction and Feature Selection

Patients undergoing CSF bacterial culture tests were identified from MIMIC-IV, yielding 11892 cases. After excluding records missing essential biochemical tests, 11859 samples with 306 features and 2 label sets were retained (Fig. 1). Structured data (n=300), including demographic, vital sign, blood test, and CSF biochemical results, underwent feature selection via the Least Absolute Shrinkage and Selection Operator (LASSO), reducing the dataset to the most relevant predictors [20]. Unstructured text data (n=6) was processed by extracting “Allergies,” “History of Present Illness,” “Past Medical History,” and “Family History” using regular expressions from discharge summaries, while "Chief Complaint” and ICD text data were directly accessed from database columns.

Data Splitting and Preprocessing

The dataset was stratified into training, validation, and test sets (6:2:2 ratio) to ensure balanced class representation [21]. Missing values were imputed using a K-Nearest Neighbors Imputer (n_neighbors = 5), and numerical features were normalized using the StandardScaler [22, 23]. Processed data was serialized for reproducibility. Text data was tokenized using BioBERT to create embeddings compatible with multimodal models.

Model Development

Three multimodal neural network architectures (Multimodal_FCN, Multimodal_LSTM, Multimodal_CNN) were developed to integrate structured and unstructured data. BioBERT embeddings of unstructured text were combined with structured data for model training: Multimodal_FCN: Unstructured data embeddings were fed into a fully connected network (FCN); Multimodal_LSTM: Embeddings were processed by a LSTM network before integration with structured data; Multimodal_CNN: Embeddings underwent a convolutional

neural network based (CNN) transformation before combining with structured data. Baseline models trained exclusively on structured data were also developed for comparison.

Data Analysis and Tools

Model performances were evaluated using accuracy, sensitivity, specificity, precision, and AUC to comprehensively assess predictive performance for binary classification and survival prediction tasks. Feature relationships were analyzed using Pearson correlation coefficients, visualized via heatmaps. Distribution differences were assessed using kernel density estimation (KDE), boxplots, and dotplots. Statistical significance was determined via two-sample t-tests. Word cloud analysis highlighted key terms in textual data, with size and color reflecting frequency and importance. Data processing and modeling were conducted using Python (v3.9.12) with libraries including pandas, numpy, and scikit-learn. TensorFlow and Keras were employed for neural network training, while BioBERT from Hugging Face handled text embedding. Due to the memory-intensive nature of the labevents file, systems with at least 64 GB RAM were required.

Results

Data Extraction and Processing

To train models for predicting CSF bacterial culture results, the MIMIC-IV dataset was refined to 9261 samples (Culture Dataset, Table 1) after sample filtering excluding incomplete data (Fig. 1). Feature selection using LASSO reduced 300 potential predictors to 27 key features (Fig. S3). This dataset comprised 9016 culture-negative and 245 culture-positive samples. For ICU survival prediction, the dataset was filtered to 5795 samples with 23 structured and 2 unstructured features (Survival Dataset, Table 2, Fig. S3), including 5500 survivors and 295 non-survivors (Fig. 1). Pre-hospital tests conducted within

seven days of admission were integrated by matching them to patient records. For example, subject_id 10003019 contributed 513 pre-hospital tests to 1771 hospital tests (Fig. S4). Across the entire cohort, the inclusion of pre-hospital data expanded the total number of tests from 7378238 to 8340248, representing an overall enhancement of approximately 13%.

Independent validation using MIMIC-III involved extracting 1618 samples for culture result prediction and 1376 for survival prediction (Fig. S5). Due to differences in available features, the culture prediction dataset included 1463 culture-negative and 155 culture-positive samples 23 features, while the survival prediction dataset included 1197 survivors and 179 non-survivors with 22 structured features and 2 text data (Figs. S6A-B).

Descriptive Analysis for Culture Dataset

Selected features included the results of CSF (C_) and blood (B_) tests, as well as vital signs, with C_polys and C_lymphs showing distinct distributions between culture-negative and culture-positive groups (Figs. 2A-D, Fig. S7). Violin and density plots revealed clustering of C_polys around 85 and C_lymphs near 5 in culture-positive cases, contrasting with 0 and 87 in culture-negative cases, respectively. A correlation heatmap (Fig. 2E) indicated low inter-feature correlations, enhancing model robustness. Table 1 summarizes baseline characteristics, showing significant differences between culture-negative and culture-positive groups. Feature importance analysis identified CSF leukocyte count, C_polys, and C_lymphs as top predictors. These findings align with their highest significance rankings based on P values when comparing culture negative and culture positive groups (Table 1).

Table 1 Baseline Characteristics of the Culture Dataset

Features	95%CI*		Sample Count		Importance**	P value#
	Negative	Positive	Negative	Positive		
C_Polys	(14.5, 15.58)	(52.74, 60.08)	8733	390	0.20065	2.56E-191
C_Lymphs	(59.01, 60.29)	(24.56, 30.72)	8733	390	0.12913	1.16E-88
C_LeukocyteCount	(32.63, 38.75)	(1281.8, 3611.6)	8730	390	0.17535	2.00E-80
C_Protein	(61.38, 68.0)	(172.1, 330.88)	8596	379	0.05021	1.42E-56
C_Glucose	(71.94, 73.04)	(55.5, 61.6)	8494	366	0.04958	1.32E-23

B_Platelet	(236.2 240.8)	(265.65, 294.27)	8810	402	0.01150	4.67E-13
B_Hemoglobin	(11.73, 11.83)	(10.81, 11.21)	8812	402	0.02114	2.24E-11
B_Albumin	(3.71, 3.77)	(3.17, 3.45)	2757	95	0.02400	1.69E-09
B_lymphocytes	(20.48, 21.3)	(12.7, 16.5)	5388	160	0.01839	2.30E-07
B_Glucose	(120.66, 122.72)	(128.64, 136.32)	8445	397	0.01080	1.22E-05
pain	(4.16, 4.52)	(4.9, 6.7)	2302	110	0.03600	0.00078
B_Potassium	(4.03, 4.05)	(3.88, 4.0)	8635	399	0.01771	0.00084
C_Macrophage	(14.57, 16.25)	(7.74, 12.58)	1876	103	0.01903	0.00480
B_ImmatureGranulocytes	(0.56, 0.62)	(0.57, 1.03)	1474	54	0.02042	0.01006
B_MonocyteCount	(0.65, 0.69)	(0.69, 0.93)	1818	69	0.05388	0.01980
B_Creatinine	(1.01, 1.05)	(0.85, 0.99)	8706	400	0.01309	0.03252
acuity	(2.09, 2.15)	(1.91, 2.13)	2781	158	0.02078	0.07788
B_Sodium	(138.83, 139.01)	(138.02, 139.06)	8636	399	0.00852	0.07870
B_I	(0.83, 0.89)	(0.6, 0.86)	2463	126	0.01553	0.11139
B_RDW-SD	(47.18, 47.78)	(47.39, 49.75)	3178	160	0.01588	0.11888
B_MCHC	(33.39, 33.45)	(33.16, 33.46)	8810	402	0.01308	0.17066
B_LactateDehydrogenase	(311.24, 352.28)	(215.98, 540.1)	2190	72	0.01713	0.43792
resprate	(17.51, 17.73)	(17.0, 18.5)	2422	131	0.02040	0.60554
B_pO2	(120.45, 130.07)	(111.06, 132.14)	1350	126	0.00881	0.65500
B_PTT	(30.19, 30.71)	(29.41, 31.79)	4989	254	0.00953	0.79900
B_MonocytesPercentage	(6.24, 6.46)	(5.65, 6.97)	5388	160	0.01411	0.91825
B_H	(24.96, 30.96)	(13.16, 42.56)	2463	126	0.00534	0.98798

* 95%CI: the 95% confidence interval of the value of each feature in culture-negative and culture-positive groups.
** Importance: relative feature importance of each feature as derived from RF algorithm.
# P value: indicator of significance levels from the T-test comparing culture-negative and culture-positive groups, sorted in descending order of significance.

Table 2 Baseline Characteristics of the Survival Dataset

Features	95%CI*		Sample Count		P value#
	Survivor	Non-Survivor	Survivor	Non-Survivor	
B_UreaNitrogen	(17.32, 18.06)	(28.43, 34.11)	4746	291	4.61E-56
B_RDW	(14.56, 14.68)	(16.43, 17.17)	4745	292	2.13E-54
B_RedBloodCells	(3.84, 3.88)	(3.16, 3.32)	4748	292	1.25E-39
B_Albumin	(3.64, 3.72)	(2.68, 2.94)	1387	99	1.67E-34
age	(54.22, 55.14)	(63.97, 67.27)	5500	295	1.47E-25
B_PT	(12.82, 12.98)	(14.06, 14.88)	2953	210	1.02E-21
B_INR(PT)	(1.15, 1.17)	(1.27, 1.35)	2953	210	3.67E-21
B_AlkalinePhosphatase	(94.85, 101.29)	(138.56, 188.72)	2090	151	1.86E-20
B_Chloride	(102.96, 103.24)	(105.04, 106.86)	4702	290	4.58E-20
B_WhiteBloodCells	(8.63, 9.05)	(11.29, 14.15)	4748	292	6.92E-17
C_Polys	(13.1, 14.42)	(20.88, 28.26)	5367	285	1.38E-12
B_PlateletCount	(237.37, 243.81)	(179.39, 208.37)	4750	292	1.25E-11
B_Calcium	(8.78, 8.82)	(8.4, 8.62)	3682	284	1.05E-10
B_MCV	(90.2, 90.58)	(92.07, 93.87)	4748	292	4.26E-10
B_LactateDehydrogenase	(279.4, 328.6)	(410.3, 756.91)	1212	110	1.85E-08
C_Protein	(63.49, 73.45)	(95.27, 161.25)	5288	281	3.26E-07
C_RBC	(2366, 3446)	(3674, 16446)	5355	275	5.04E-07

acuity	(2.15, 2.21)	(1.76, 2.0)	2722	112	3.45E-06
C_LactateDehydrogenase	(30.05, 36.87)	(22.31, 106.21)	865	68	0.000366271
resprate	(17.51, 17.71)	(17.63, 19.15)	2464	92	0.005717448
B_Eosinophils	(1.48, 1.74)	(0.6, 1.48)	2327	110	0.059304815
B_Phosphate	(3.4, 3.46)	(3.2, 3.52)	3660	284	0.2089438
C_Eosinophils	(2.52, 3.78)	(2.09, 5.69)	550	36	0.560681978

* 95%CI: the 95% confidence interval of the value of each feature in ICU survivor and non-survivor groups.
# P value: indicator of significance levels from the T-test comparing ICU survivor and non-survivor groups, sorted in descending order of significance.

Culture Result Predicting Model Development

Models (CNN, FCN, RF) were trained on 9261 Culture Dataset while CNN performed best, achieving an AUC of 0.813, accuracy of 0.837, and sensitivity of 0.735 in internal validation, with slightly lower metrics on validation dataset (AUC 0.769, accuracy 0.781, and sensitivity 0.729) (Table 3, Fig. 3A-B). To improve the accuracy of diagnostic labeling, culture-positive samples were redefined following clinical evidence to include cases with elevated CSF leukocytes (>1000 cells/μL), high polymorphonuclear leukocyte percentages (>50%), or clinical chief complaint mentions of "bacterial meningitis" as culture-positive classification [24, 25]. This adjustment increased culture-positive samples from 245 to 404 and reduced culture-negative samples 9016 to 8857. Re-trained models on corrected labels demonstrated improved performance. The FCN model achieved the highest AUC of 0.891, accuracy of 0.867, precision of 0.244 and sensitivity of 0.833 (Table 3, Fig. 3C) and consistent performance trends was observed in independent validation dataset (Fig. 3D).

Table 3 Performance Evaluation for Culture Result Predicting Models

Prediction Label	Datasets	Methods	Accuracy	Sensitivity	Specificity	Precision	AUC
Culture Result	Internal Validation	**CNN**	**0.837**	**0.735**	**0.840**	**0.111**	**0.813**
		FCN	0.834	0.714	0.837	0.106	0.809
		RF	0.823	0.694	0.826	0.098	0.804
	Independent Validation	CNN	0.781	0.729	0.784	0.198	0.769
Redefined Culture Result	Internal Validation	**FCN**	**0.867**	**0.833**	**0.868**	**0.244**	**0.891**
		CNN	0.855	0.822	0.857	0.227	0.876
		RF	0.847	0.741	0.852	0.186	0.830
	Independent Validation	FCN	0.815	0.794	0.817	0.315	0.853

Descriptive Analysis of Survival Dataset

From the filtered 23 structured features, the correlation heatmap revealed low inter-feature redundancy (Fig. 4A). Analysis of unstructured data revealed key differences in clinical text between survivors and non-survivors. Chief complaint terms like "headache" and "weakness" were frequent in survivors, while terms like "mental" and "altered" predominated among non-survivors (Figs. 4B-C). ICD text analysis highlighted “history” and “personal” in survivors, compared to “acute” and “failure” in non-survivors (Figs. 4D-E). To optimize computational efficiency, the token counts for Chief Complaint and ICD texts were truncated to 20 and 379 tokens, respectively, preserving 95% of relevant information (Figs. 4F-G). Additional distributions and adjustments for other clinical text categories, such as Family History and Past Medical History, are detailed in Fig. S8-S9, with word clouds provided in Fig. S10. Table S2 summarizes the length distributions for each unstructured data type, and Table S3 details the token lengths following BioBERT transformation.

Comparison of structured and unstructured data characteristics across the Survival Dataset and matched validation datasets showed consistency, providing reliable inputs for model development (Fig. S6B, word frequency in Fig. S9, word clouds in Fig. S10).

Model Development and Validation for Survival Prediction

Three neural network architectures (Multimodal_FCN, Multimodal_LSTM, and Multimodal_CNN) were designed to integrate structured and unstructured data (Fig. 5A, Fig. S12). Fine-tuned BioBERT adapted to clinical text, optimizing tokenization for Chief Complaint (CC) and ICD texts as well as other text features (Table S3). A baseline model trained solely on structured data achieved an AUC of 0.809, accuracy of 0.745, precision of 0.131, and sensitivity of 0.712 (Table S4). Incorporating unstructured data significantly improved performance. Leveraging both structured and optimized unstructured features (CC

and ICD text), Multimodal_CNN achieved the best metrics: AUC of 0.933, accuracy of 0.850, precision of 0.233 and sensitivity of 0.847 (Table 4; Fig. 5B-D). Independent validation with MIMIC-III data showed Multimodal_CNN achieved an AUC of 0.889, accuracy of 0.814, precision of 0.397 and sensitivity of 0.827 (Fig. 5E; Table S5).

Table 4 Comparing the Predictive Performance of Three Multimodal Models

Model	Feature Subset	Accuracy	Sensitivity	Specificity	Precision	AUC	AUC Diff*
Multimodal_FCN	All Dataset	0.821	0.814	0.822	0.197	0.894	baseline
	Drop Chief Complaint	0.801	0.780	0.802	0.174	0.883	-0.011
	Drop ICD	0.768	0.763	0.768	0.150	0.863	-0.031
	Drop Ill History	0.821	0.814	0.821	0.196	0.910	0.016
	Drop Medical History	0.830	0.814	0.831	0.205	0.895	0.001
	Drop Allergies	0.829	0.831	0.829	0.207	0.904	0.01
	Drop Family History	0.847	0.814	0.849	0.224	0.896	0.002
	Structured Data + Chief Complaint + ICD	0.828	0.831	0.828	0.206	0.900	0.006
Multimodal_LSTM	All Dataset	0.808	0.763	0.810	0.177	0.857	baseline
	Drop Chief Complaint	0.770	0.746	0.772	0.149	0.856	-0.001
	Drop ICD	0.765	0.763	0.765	0.149	0.852	-0.005
	Drop Ill History	0.777	0.780	0.776	0.158	0.863	0.006
	Drop Medical History	0.801	0.763	0.803	0.172	0.863	0.006
	Drop Allergies	0.793	0.763	0.795	0.166	0.861	0.004
	Drop Family History	0.790	0.763	0.792	0.164	0.865	0.008
	Structured Data + Chief Complaint + ICD	0.798	0.763	0.800	0.170	0.857	0
Multimodal_CNN	All Dataset	0.850	0.847	0.850	0.233	0.932	baseline
	Drop Chief Complaint	0.827	0.831	0.827	0.205	0.918	-0.014
	Drop ICD	0.778	0.763	0.779	0.156	0.869	-0.063
	Drop Ill History	0.843	0.847	0.843	0.224	0.927	-0.005
	Drop Medical History	0.846	0.847	0.845	0.227	0.924	-0.008
	Drop Allergies	0.849	0.847	0.849	0.231	0.934	0.002
	Drop Family History	0.848	0.847	0.848	0.230	0.930	-0.002
	Structured Data + Chief Complaint + ICD	**0.850**	**0.847**	**0.850**	**0.233**	**0.933**	**0.001**

*AUC diff: difference of AUC between baseline model (with all 6 text data) and modified models (with 1 text data dropped).

Discussion

This study aimed to create a patient-centered approach that combines standard lab test results with clinical evidence to enable culture-independent diagnosis and facilitate early

management of suspected meningitis. To achieve this, machine learning algorithms were utilized to train, test, and validate datasets of ICU patients from the MIMIC Clinical Database. First, a data mining and classification algorithm (FCN) was refined to predict the likelihood of a positive CSF culture result, thereby supporting the diagnosis of BM without solely relying on traditional culture methods. Next, a predictive multimodal model (Multimodal_CNN) was enhanced to forecast ICU survival by integrating both structured and unstructured data, demonstrating strong potential to improve the identification of high-risk patients and support early intervention strategies.

A standard CSF culture typically takes 1 to 3 days to grow and identify bacteria under lab conditions [26]. For suspected bacterial infections especially acute BM, an empiric antibiotic treatment often begins before the CSF culture results are available. Clinicians are increasingly relying on molecular tests like Polymerase Chain Reaction (PCR) for identifying pathogens due to PCR's speed and accuracy in detecting bacterial DNA [27, 28]. However, the application of PCR can be constrained by issues such as availability, cost, false positives, and limited pathogen detection [29, 30]. Here our study demonstrated the effectiveness of CNN algorithms in utilizing regular blood and CSF test result and accurately predicting positivity of CSF culture results, achieving an AUC of 0.813 and an accuracy rate of 83.7% in validation dataset. This result shows promise of the CNN model as a valuable tool in assisting clinical decision-making with a critical advantage by providing results instantly, bypassing the typical 24 to 72-hour delay as required for traditional CSF culture.

Studies have shown that CSF culture may miss bacterial cases in up to 40-60% of cases if antibiotics were administered before the lumbar puncture [31, 32]. Multiple studies comparing bacterial culture with PCR have proved 30–50% of CSF specimens that are negative on culture test positive with PCR [33]. Observing the limitations of culture results alone, we modified the machine learning input data in our BM diagnosis model training to

use corrected culture labels that integrate additional clinical evidence. This evidence included CSF markers, patient symptoms, and ICD codes relevant to meningitis-related diagnoses. This change increased the positive case count from 247 to 404 cases out of 9261, leading to improved training results. Notably, the model's AUC score improved from 0.813 to 0.891 during the training phase after integrating this additional clinical information, highlighting a significant boost in predictive accuracy and robustness for clinical application.

MIMIC-IV, as an evolution of MIMIC-III, has been widely used for developing and testing complex modeling machine learning models in healthcare research [15, 18]. The key improvements of MIMIC-IV include enhanced details on diagnoses and procedures using ICD-10 codes (where MIMIC-III used ICD-9 codes), and more refined dataset with better support for unstructured data processing through tools like NLP. MIMIC-IV has been extensively employed in mortality prediction and survival analysis. Classical machine learning models such as XGBoost and logistic regression often achieve robust AUC scores around 0.80 to 0.85 [34-36]. This level of performance makes these models strong candidates for applications in clinical prediction tasks, especially when interpretability and computational efficiency are key requirements. Taking advantage of the updated and additional data integrated in MIMIC-IV, the hypothesis of this study was that multimodal machine learning model can deliver precise survival prediction and personalized patient management. BERT and its model variants (BioBERT, Bio_ClinicalBERT etc.) have been widely used to extract textual features and their relationships from free-text biomedical data [11, 12, 37]. BioBERT, an extension of BERT, was pre-trained on PubMed abstracts and full-text articles from PubMed Central, both of which are databases rich in biomedical literature [11, 12, 37]. Bio_ClinicalBERT, on the other hand, was trained on clinical notes from the MIMIC-III dataset, making it more specialized for clinical text. However, its application in large urgent care datasets remains limited. In our study, independent validation

was conducted using data from the MIMIC-III database. To prevent potential information leakage, we opted to use BioBERT for training instead of Bio_ClinicalBERT. Interestingly, when we tested both BioBERT and Bio_ClinicalBERT for converting chief complaints and other textual data into feature vectors, we found that the vectors generated by the two models were identical. This finding suggests that despite their different training corpora, these models can yield consistent representations for certain types of clinical text. A BioBERT-based model was developed to accurately predict patient mortality risk using a variety of time-series data and clinical records from the MIMIC-III database [38]. Here our testing and validation results demonstrated that integrating BioBERT to process unstructured data, such as chief complaints from the MIMIC-IV dataset, significantly enhanced the predictive accuracy of CNN and RNN models. The integration led to the best model performance, achieving an AUC score exceeding 0.85. This improvement in the present study highlights the potential of using advanced NLP techniques and datasets to better interpret clinical narratives and enhance predictive modeling in critical care settings.

Together, this study highlights the potential of an integrated machine learning-based system to support patient management and improve clinical decision-making. For instance, such a system could include tools like a Clinical Decision Support System (CDSS) to operationalize these findings in clinical settings [39]. We envision deploying the culture-free BM diagnostic algorithm within a broader predictive framework that continuously monitors patients with suspected meningitis and functions as an early warning system, identifying those at risk for BM even before culture results are available. By incorporating survival prediction capabilities, the system can provide real-time, personalized risk assessments, enabling clinicians to identify patients at higher risk for severe outcomes. Through the integration of structured and unstructured data processed with advanced machine learning techniques such as CNNs and NLP, this multimodal approach can deliver precise survival forecasts and support care

prioritization. By offering this predictive assistance, the system ensures timely interventions, reduces the risk of under-treatment or oversight, and empowers clinicians to make more informed decisions, even in high-pressure environments with heavy patient loads.

A potential limitation of both the diagnostic and survival prediction models is their lower recall metric, which indicates a tendency toward high false positive rates. In the context of the diagnostic model for BM, a false positive could lead to unnecessary or prolonged antibiotic treatment. This is rather a conservative and minor error as the current clinical guidelines generally recommend that if BM is suspected, antibiotics should be administered within an hour of the patient presenting with symptoms [3, 40]. Studies indicate that over 90% of suspected BM cases are treated empirically with antibiotics immediately after or even before lumbar puncture [41]. In survival prediction, particularly in critical conditions like BM, a lower precision can sometimes be acceptable, as it results in more conservative treatment, which is generally safer in these high-risk scenarios. Although these interventions, such as closer monitoring and early treatment, are not expected to result in harm but over-prioritizing these patients may strain healthcare resources in high-demand ICU settings. Despite the models achieving solid metrics in terms of sensitivity, specificity, and AUC, their relatively lower recall highlights the need for further optimization. To mitigate the challenge of data imbalance, we explored various strategies, including oversampling, undersampling, and back-translation to augment the minority class in textual data. Despite these efforts, none of these approaches yielded a measurable enhancement in the model's predictive performance. This remains the case even after applying various approaches to mitigate the effects of class imbalance in the training dataset, which can otherwise skew predictions toward the majority class and reduce recall for minority class outcomes.

Conclusion

This study demonstrates the potential of multimodal machine learning models, including CNN and FCN, for predicting CSF culture results to support BM diagnosis and ICU survival outcomes. Using structured and unstructured data from the MIMIC-IV database, these models achieved strong performance in identifying BM and forecasting survival. Integrated into clinical decision-making frameworks, such tools could provide rapid, data-driven insights to enhance diagnosis, intervention, and patient outcomes in ICU settings.

Abbreviations

AUC	Area Under the ROC Curve
BERT	Bidirectional Encoder Representations from Transformers
BioBERT	BERT pretrained on biomedical data
BM	Bacterial meningitis
CC	Chief Complaint
CDSS	Clinical Decision Support System
CNN	Convolutional Neural Network
CSF	Cerebrospinal fluid
FCN	Fully Connected Network
ICD	International Classification of Diseases
ICU	Intensive Care Unit
KDE	Kernel density estimation
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long short-term memory
MIMIC	Medical Information Mart for Intensive Care
NLP	Natural Language Processing
PCR	Polymerase Chain Reaction

RF Random Forest

Supplementary Information

Additional file 1: Figure S1. Workflow and Example of the Word2Vec Embedding Process using BioBERT. **Figure S2.** Workflow for Chief Complaint Extraction from Clinical Text. **Figure S3.** Feature Selection with LASSO. **Figure S4.** Total Diagnostic Tests Pre- and Post-Admission for a Representative Patient (10003019). **Figure S5.** Flowchart for the Extraction of the Independent Validation Dataset. **Figure S6.** Correlation Heatmap of the Features in the Independent Validation Dataset. **Figure S7.** Distribution of Culture Dataset Features. **Figure S8.** Cumulative Frequency of Text Lengths for Key Text Data. **Figure S9.** Cumulative Probability Plots of Token Counts in Unstructured Data Samples Transformed by BioBERT. **Figure S10.** Word Clouds Generated from the Textual Data. **Figure S11.** Top 10 Words Ranked by Frequency Across Various Text Categories. **Figure S12.** CNN Architecture. **Table S1.** List of Features and Their Abbreviations. **Table S2.** String Lengths in Each Unstructured Data Type. **Table S3.** Token Counts in Unstructured Data After BioBERT Word2Vec Transformation. **Table S4.** Performance Evaluation of CNN Model Trained on Structured Data Only for ICU Outcome Prediction. **Table S5.** Performance Evaluation of Multimodal_CNN Model for ICU Outcome Prediction on Independent Dataset.

Acknowledgments

We extend our gratitude to the MIT Laboratory for Computational Physiology and the MIMIC data curators for creating and maintaining the MIMIC-III and MIMIC-IV critical care databases. We also acknowledge the patients whose data contributed to these datasets, enabling this research. Without their generous participation and the efforts of those involved in the MIMIC projects, our study would not have been possible.

Author contributions

XZ designed the workflow, drafted initial analysis code, finalized the manuscript, and

supervised all aspects of the study. RC and CY (co-first authors) built the data-analysis pipeline, performed experiments, analyzed and illustrated the data, and prepared the initial manuscript draft. SZ contributed to data interpretation and assisted in data visualization. ZH, MS, WL and DY aided in experimental design and reviewed the manuscript.

Funding

This work was supported by National Natural Science Foundation of China (81672627; 82071863).

Availability of data and materials

The data utilized in this study are derived from the MIMIC-III (https://doi.org/10.13026/8a4q-w170) and MIMIC-IV (https://doi.org/10.13026/a3wn-hq05) critical care databases, which are accessible to researchers who fulfill the credentialing requirements set by the data custodians. Due to these restrictions, the raw data are not publicly available. However, upon obtaining the necessary permissions from the MIMIC data providers, the data can be accessed for replication or further study.

The complete set of code for implementing the machine learning models, including those predicting CSF culture results and ICU outcomes, is openly available in our GitHub repository: https://github.com/AaronChen007/CNS_Model_Code.

Declarations

Ethics approval and consent to participate

The research utilized the MIMIC database, which was initially approved for use by the Institutional Review Boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology. Since MIMIC is a publicly available, de-identified dataset, individual patient consent was not required. All analyses and research activities strictly adhered to the MIMIC data access agreement and data usage guidelines.

Consent for publication

All authors have read the manuscript and are consentaneous for publication.

Competing interests

The authors have declared that no conflict of interest exists.

Reference

1. Dugue R, Kim C, Boruah A, Harrigan E, Sun Y, Thakur KT: **Time to Confirmed Neuroinfectious Diagnoses: Diagnostic Testing and Resource Allocation.** *Neurohospitalist* 2024, **14:**296-300.
2. Ippolito M, Giarratano A, Cortegiani A: **Healthcare-associated central nervous system infections.** *Curr Opin Anaesthesiol* 2022, **35:**549-554.
3. Posadas E, Fisher J: **Pediatric bacterial meningitis: an update on early identification and management.** *Pediatr Emerg Med Pract* 2018, **15:**1-20.
4. Giuliano S, Flammini S, Zanus-Fortes A, Montanari L, Leanza GM, Attanasio V, Pecori D, Sbrana F, Tascini C: **Pragmatic overview on acute bacterial and fungal infections of the central nervous system: a holistic update from diagnosis to treatment.** *New Microbiol* 2023, **46:**95-119.
5. Gomes HR: **Cerebrospinal fluid analysis: current diagnostic methods in central nervous system infectious diseases.** *Arq Neuropsiquiatr* 2022, **80:**290-295.
6. Tunkel AR, Hasbun R, Bhimraj A, Byers K, Kaplan SL, Scheld WM, van de Beek D, Bleck TP, Garton HJL, Zunt JR: **2017 Infectious Diseases Society of America's Clinical Practice Guidelines for Healthcare-Associated Ventriculitis and Meningitis.** *Clin Infect Dis* 2017, **64:**e34-e65.
7. Solomon IH: **Molecular and Histologic Diagnosis of Central Nervous System Infections.** *Surg Pathol Clin* 2020, **13:**277-289.
8. Wall EC, Chan JM, Gil E, Heyderman RS: **Acute bacterial meningitis.** *Curr Opin Neurol* 2021, **34:**386-395.
9. Rai S, Mishra P, Ghoshal UC: **Survival analysis: A primer for the clinician scientists.** *Indian J Gastroenterol* 2021, **40:**541-549.
10. Chiu CC, Wu CM, Chien TN, Kao LJ, Li C, Chu CM: **Integrating Structured and Unstructured EHR Data for Predicting Mortality by Machine Learning and Latent Dirichlet Allocation Method.** *Int J Environ Res Public Health* 2023, **20**.
11. Mansour M, Yilmaz F, Miletic M, Sariyar M: **What Kind of Transformer Models to Use for the ICD-10 Codes Classification Task.** *Stud Health Technol Inform* 2024, **316:**1008-1012.
12. Alsentzer E, Murphy JR, Boag W, Weng W-H, Jin D, Naumann T, McDermott M: **Publicly available clinical BERT embeddings.** *arXiv preprint arXiv:190403323* 2019.
13. Huang K, Gray TF, Romero-Brufau S, Tulsky JA, Lindvall C: **Using nursing notes to improve clinical outcome prediction in intensive care patients: A retrospective cohort study.** *Journal of the American Medical Informatics Association* 2021, **28:**1660-1666.
14. Yu K, Zhang M, Cui T, Hauskrecht M: **Monitoring ICU mortality risk with a long short-term memory recurrent neural network.** In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020*. World Scientific; 2019: 103-114.
15. Johnson AE, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Hao S, Moody B, Gow B: **MIMIC-IV, a freely accessible electronic health record dataset.** *Scientific data* 2023, **10:**1.

451 16. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody
452 GB, Peng CK, Stanley HE: **PhysioBank, PhysioToolkit, and PhysioNet: components of a new**
453 **research resource for complex physiologic signals.** *Circulation* 2000, **101:**E215-220.
454 17. Johnson AE, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, Moody B, Szolovits P,
455 Anthony Celi L, Mark RG: **MIMIC-III, a freely accessible critical care database.** *Scientific data*
456 2016, **3:**1-9.
457 18. Johnson A, Pollard, T., & Mark, R.: **MIMIC-III Clinical Database CareVue subset (version 1.4).**
458 *PhysioNet https://doiorg/1013026/8a4q-w170* 2022.
459 19. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm
460 E, Langan SM, Committee RW: **The REporting of studies Conducted using Observational**
461 **Routinely-collected health Data (RECORD) statement.** *PLoS medicine* 2015, **12:**e1001885.
462 20. Ranstam J, Cook JA: **LASSO regression.** *British Journal of Surgery* 2018, **105:**1348-1348.
463 21. Singh R, Mangat NS, Singh R, Mangat NS: **Stratified sampling.** *Elements of survey sampling*
464 1996:102-144.
465 22. Beretta L, Santaniello A: **Nearest neighbor imputation algorithms: a critical evaluation.** *BMC*
466 *Med Inform Decis Mak* 2016, **16 Suppl 3:**74.
467 23. Ali PJM, Faraj RH, Koya E, Ali PJM, Faraj RH: **Data normalization and standardization: a**
468 **technical report.** *Mach Learn Tech Rep* 2014, **1:**1-6.
469 24. Leib SL, Boscacci R, Gratzl O, Zimmerli W: **Predictive value of cerebrospinal fluid (CSF)**
470 **lactate level versus CSF/blood glucose ratio for the diagnosis of bacterial meningitis following**
471 **neurosurgery.** *Clinical infectious diseases* 1999, **29:**69-74.
472 25. Ortiz OHH, García HIG, Ramírez FM, Flórez JSC, Valencia BAG, Mantilla SEM, Ochoa MJM,
473 Ochoa JES, Jaimes F: **Development of a prediction rule for diagnosing postoperative**
474 **meningitis: a cross-sectional study.** *Journal of neurosurgery* 2017, **128:**262-271.
475 26. Trujillo-Gómez J, Navarro CE, Atehortúa-Muñoz S, Florez ID: **Acute infections of the central**
476 **nervous system in children and adults: diagnosis and management.** *Minerva medica* 2024,
477 **115:**476-502.
478 27. Ghoweba Y, Shabestari SAS, Malik ZA: **Diagnostic Accuracy of Cerebrospinal Fluid Multiplex**
479 **Polymerase Chain Reaction Panel Testing in Patients With Suspected Central Nervous**
480 **System Infections: A Multi-Center Study in the United Arab Emirates.** *Cureus* 2024, **16**.
481 28. Sunnerhagen T, Widén J, Handhal S, Özkaya Şahin G: **A retrospective observational study of**
482 **1000 consecutive patients tested with the FilmArray® Meningitis/Encephalitis panel: clinical**
483 **diagnosis at discharge and microbiological findings.** *Sci Rep* 2024, **14:**4015.
484 29. Obaro S: **Updating the diagnosis of bacterial meningitis.** *Lancet Infect Dis* 2019,
485 **19:**1160-1161.
486 30. Young N, Thomas M: **Meningitis in adults: diagnosis and management.** *Intern Med J* 2018,
487 **48:**1294-1307.
488 31. Michael B, Menezes BF, Cunniffe J, Miller A, Kneen R, Francis G, Beeching NJ, Solomon T:
489 **Effect of delayed lumbar punctures on the diagnosis of acute bacterial meningitis in adults.**
490 *Emerg Med J* 2010, **27:**433-438.

32. Waldrop G, Zucker J, Boubour A, Radmard S, Green DA, Thakur KT: **Clinical Significance of Positive Results of the BioFire Cerebrospinal Fluid FilmArray Meningitis/Encephalitis Panel at a Tertiary Medical Center in the United States.** *Arch Pathol Lab Med* 2022, **146:**194-200.
33. Pallerla SR, Van Dong D, Linh LTK, Van Son T, Quyen DT, Hoan PQ, Trung NT, The NT, Rüter J, Boutin S, et al: **Diagnosis of pathogens causing bacterial meningitis using Nanopore sequencing in a resource-limited setting.** *Ann Clin Microbiol Antimicrob* 2022, **21:**39.
34. Lin S, Lu W, Wang T, Wang Y, Leng X, Chi L, Jin P, Bian J: **Predictive model of acute kidney injury in critically ill patients with acute pancreatitis: a machine learning approach using the MIMIC-IV database.** *Ren Fail* 2024, **46:**2303395.
35. Xie W, Li Y, Meng X, Zhao M: **Machine learning prediction models and nomogram to predict the risk of in-hospital death for severe DKA: A clinical study based on MIMIC-IV, eICU databases, and a college hospital ICU.** *Int J Med Inform* 2023, **174:**105049.
36. Hou N, Li M, He L, Xie B, Wang L, Zhang R, Yu Y, Sun X, Pan Z, Wang K: **Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost.** *J Transl Med* 2020, **18:**462.
37. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J: **BioBERT: a pre-trained biomedical language representation model for biomedical text mining.** *Bioinformatics* 2020, **36:**1234-1240.
38. Niu K, Zhang K, Peng X, Pan Y, Xiao N: **Deep multi-modal intermediate fusion of clinical record and time series data in mortality prediction.** *Front Mol Biosci* 2023, **10:**1136071.
39. Chen K, Luo L, Tan Y, Chen G: **Medical diagnosis based on artificial intelligence and decision support system in the management of health development.** *J Eval Clin Pract* 2024.
40. Tunkel AR, Hartman BJ, Kaplan SL, Kaufman BA, Roos KL, Scheld WM, Whitley RJ: **Practice guidelines for the management of bacterial meningitis.** *Clin Infect Dis* 2004, **39:**1267-1284.
41. Mount HR, Boyle SD: **Aseptic and Bacterial Meningitis: Evaluation, Treatment, and Prevention.** *Am Fam Physician* 2017, **96:**314-322.

Figure Legends:

Figure 1. Overview of Data Processing and Feature Selection. Flowchart illustrating data extraction from the MIMIC-IV database, followed by filtering to ensure dataset integrity. Data was divided into two subsets (Culture Dataset and Survival Dataset), each undergoing tailored feature selection to identify critical variables for analysis and model development.

Figure 2. Descriptive Analysis of Culture Dataset. (A, B) Violin and box plots showing the distribution of the count of CSF polymorphonuclear cells (C_polys) and CSF lymphocytes (C_lymphs) in culture-positive ("Pos") and culture-negative ("Neg") groups. (C, D)

Probability density plots illustrating feature distributions as indicated. (E) Heatmap depicting correlations among individual features.

Figure 3. Evaluation of Culture Prediction Models Before and After Label Correction.

The area under the ROC (Receiver Operating Characteristic) curve (AUC) quantifies model performance. (A, C) ROC curves for culture result models evaluated using the MIMIC-IV internal validation dataset, with (A) showing results before label correction and (C) after label correction. (B, D) ROC curves for the best-performing model from the internal validation, evaluated on the MIMIC-III independent dataset, with (B) showing results before label correction and (D) after label correction.

Figure 4. Descriptive Analysis of Survival Dataset. (A) Heatmap illustrating correlations among features in structured data. (B–E) Top 10 most frequent words across various text categories; see Fig. S11 for detailed analysis. (F, G) Cumulative probability plots of token counts for unstructured data samples as processed by BioBERT.

Figure 5. Evaluation of Predictive Models for ICU Survival. (A) Schematic representation of three integrated neural network architectures (multimodal models) trained by combining structured and unstructured data. (B–D) AUC curves for models trained using multimodal models. "All" denotes models trained with structured data and all six text types. "Drop specific data type" indicates exclusion of one text type from "All" ("CC" = Chief Complaint, "IH" = Illness History, "MH" = Medical History, "AH" = Allergies History, "FH" = Family History, "ICD"). "Final" represents models trained with structured data, chief complaints, and ICD text only. (E) Model performance evaluation on an independent dataset. In panel (D), the black line shows performance using only structured data. Detailed architecture is available in Fig. S12.

Figure 1

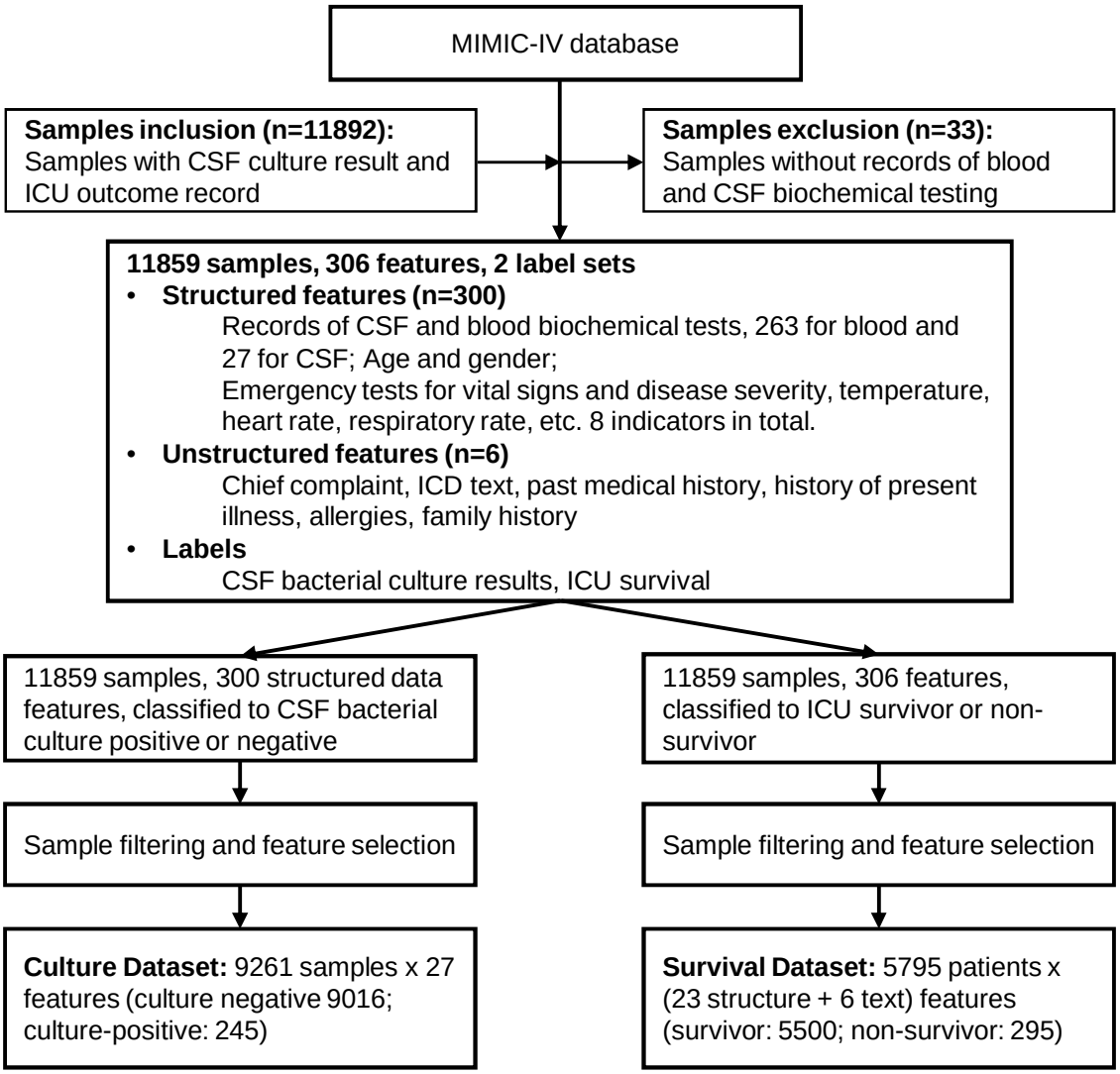

MIMIC-IV database
Samples inclusion (n=11892):
Samples with CSF culture result and ICU outcome record
Samples exclusion (n=33):
Samples without records of blood and CSF biochemical testing
11859 samples, 306 features, 2 label sets
• Structured features (n=300)
Records of CSF and blood biochemical tests, 263 for blood and 27 for CSF; Age and gender;
Emergency tests for vital signs and disease severity, temperature, heart rate, respiratory rate, etc. 8 indicators in total.
• Unstructured features (n=6)
Chief complaint, ICD text, past medical history, history of present illness, allergies, family history
• Labels
CSF bacterial culture results, ICU survival
11859 samples, 300 structured data features, classified to CSF bacterial culture positive or negative
11859 samples, 306 features, classified to ICU survivor or non-survivor
Sample filtering and feature selection
Sample filtering and feature selection
Culture Dataset: 9261 samples x 27 features (culture negative 9016; culture-positive: 245)
Survival Dataset: 5795 patients x (23 structure + 6 text) features (survivor: 5500; non-survivor: 295)

Figure 2

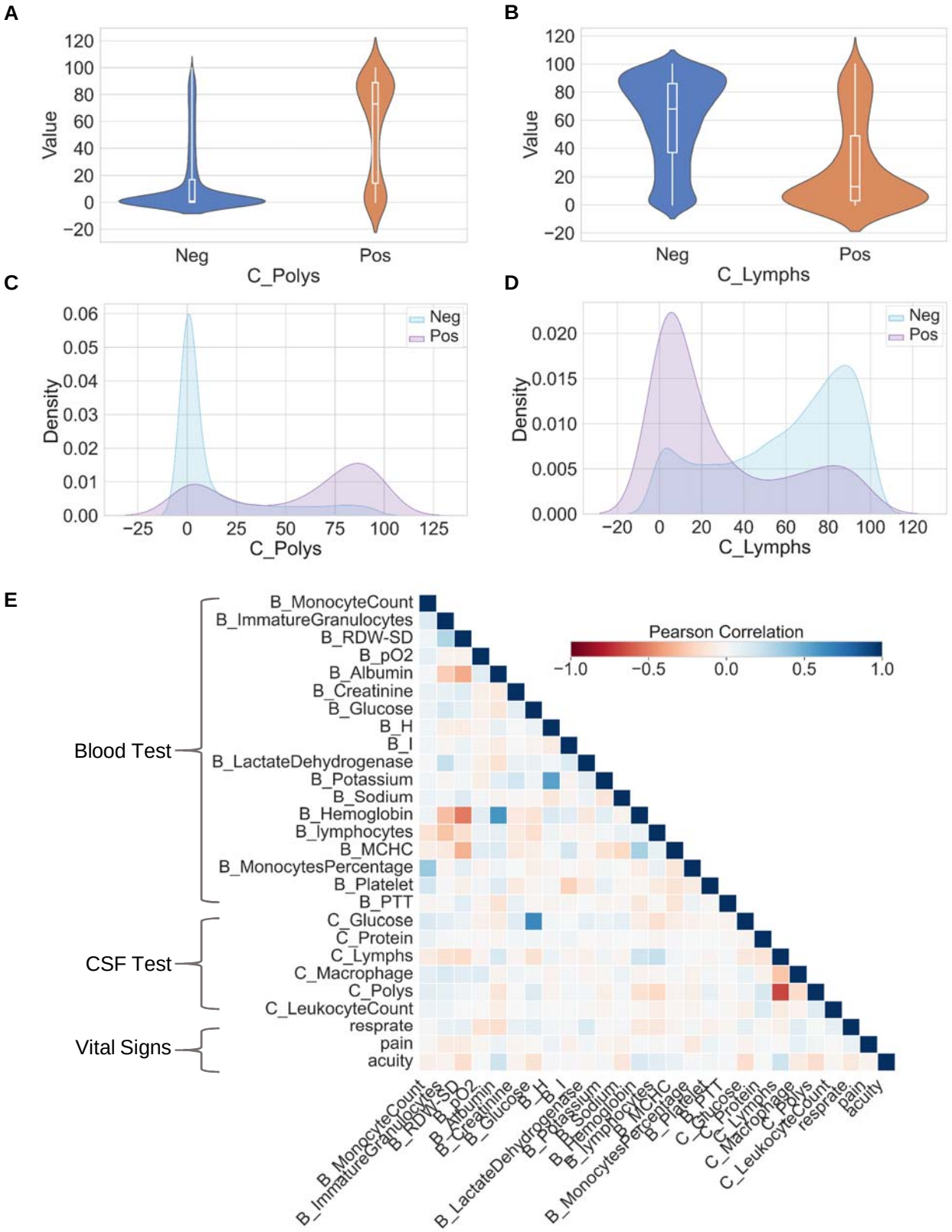

A
Value
Neg
Pos
C_Polys
B
Value
Neg
Pos
C_Lymphs
C
Density
Neg
Pos
C_Polys
D
Density
Neg
Pos
C_Lymphs
E
Pearson Correlation
Blood Test
CSF Test
Vital Signs
B_MonocyteCount
B_ImmatureGranulocytes
B_RDW-SD
B_pO2
B_Albumin
B_Creatinine
B_Glucose
B_H
B_I
B_LactateDehydrogenase
B_Potassium
B_Sodium
B_Hemoglobin
B_lymphocytes
B_MCHC
B_MonocytesPercentage
B_Platelet
B_PTT
C_Glucose
C_Protein
C_Lymphs
C_Macrophage
C_Polys
C_LeukocyteCount
resprate
pain
acuity

Figure 3

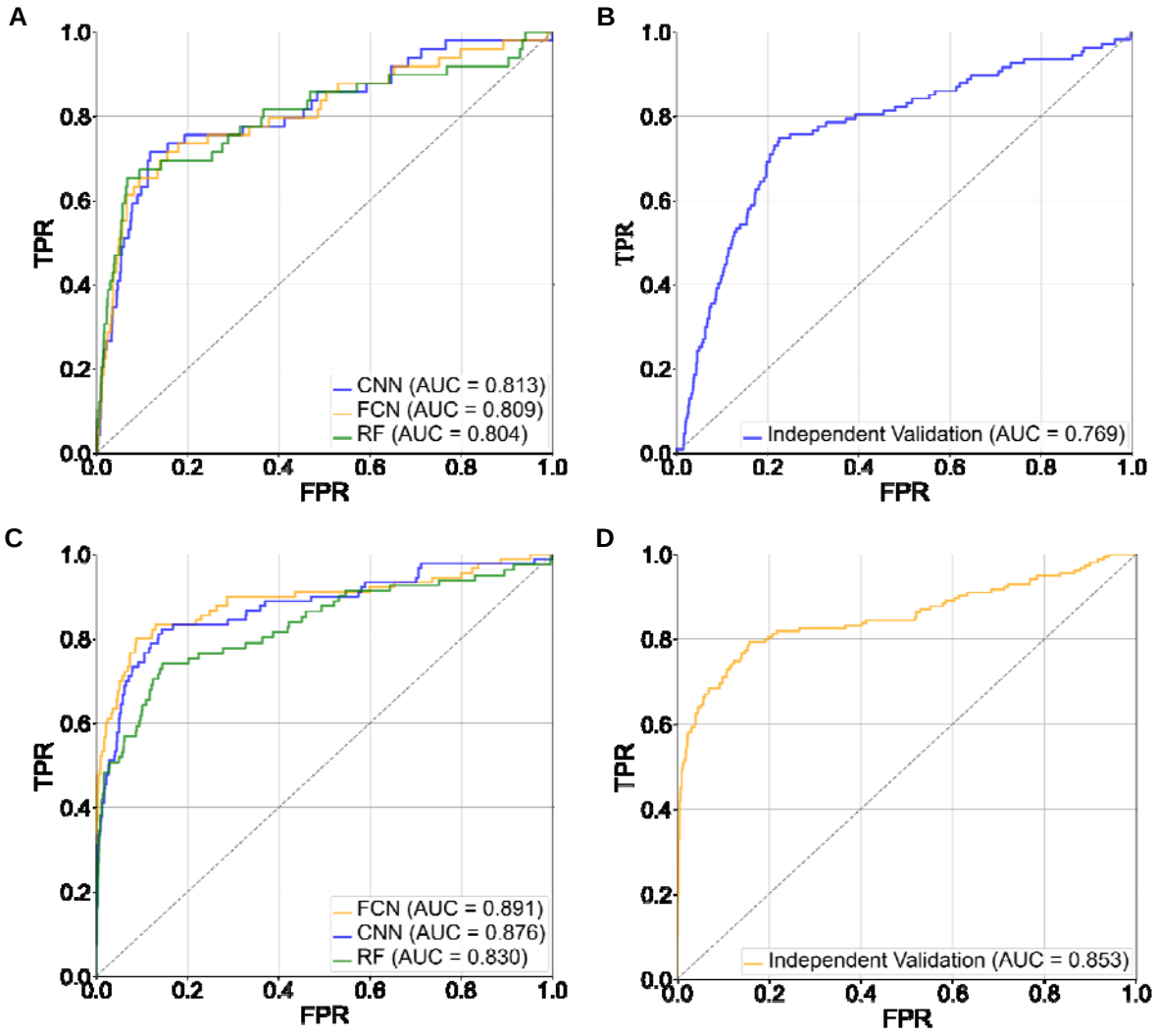

A
TPR
FPR
CNN (AUC = 0.813)
FCN (AUC = 0.809)
RF (AUC = 0.804)
B
TPR
FPR
Independent Validation (AUC = 0.769)
C
TPR
FPR
FCN (AUC = 0.891)
CNN (AUC = 0.876)
RF (AUC = 0.830)
D
TPR
FPR
Independent Validation (AUC = 0.853)

Figure 4

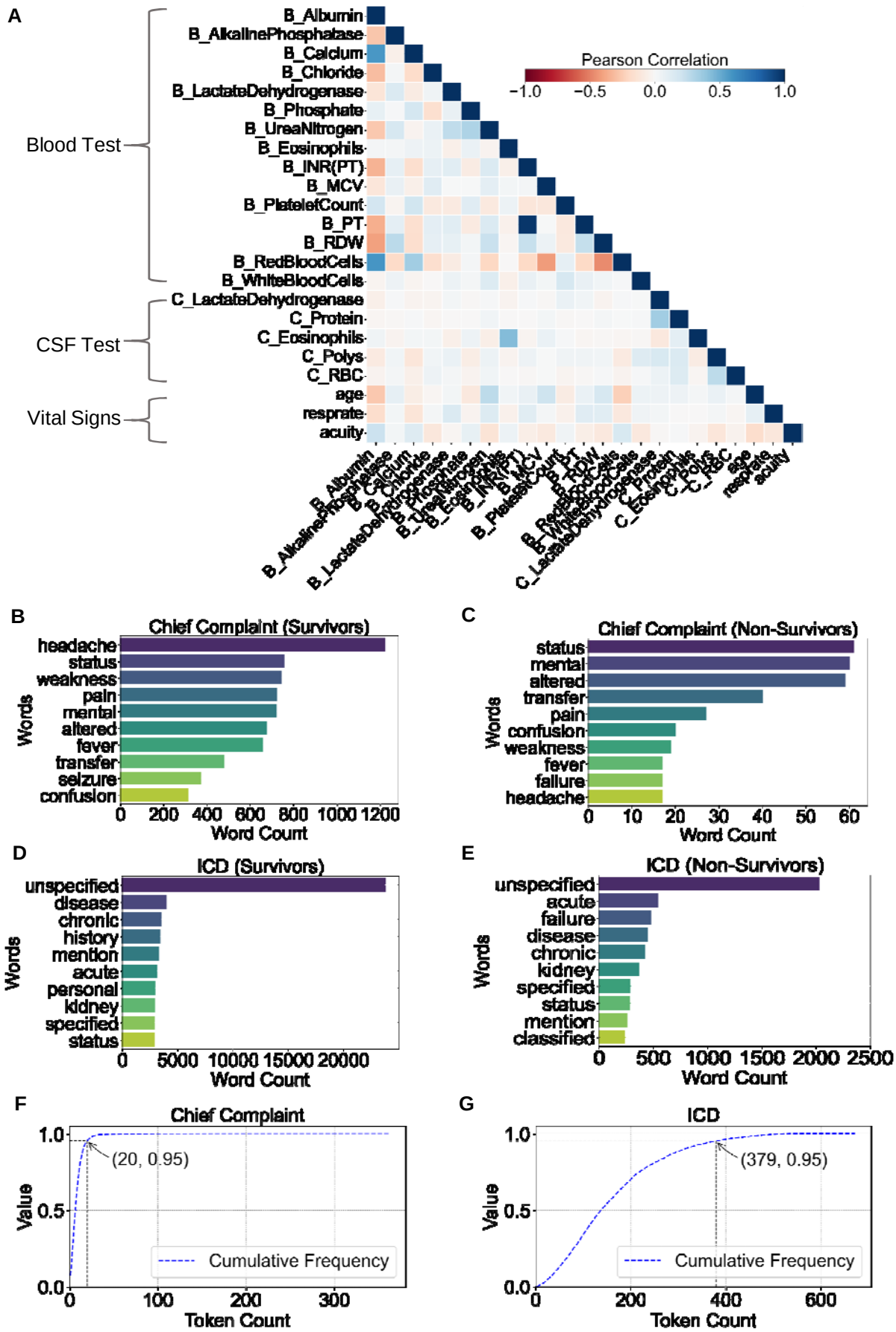
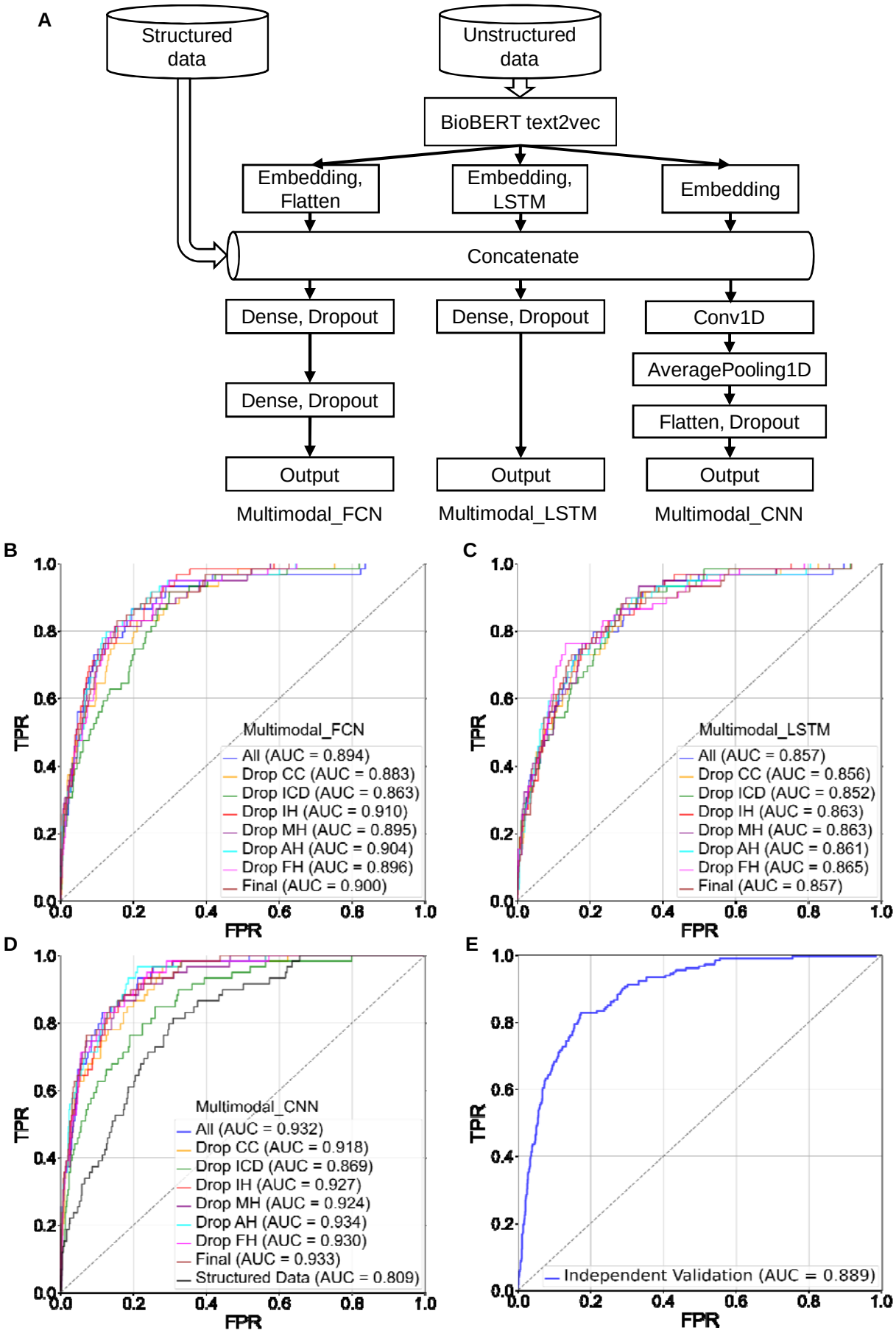

A
Blood Test
CSF Test
Vital Signs
B_Albumin
B_AlkalinePhosphatase
B_Calcium
B_Chloride
B_LactateDehydrogenase
B_Phosphate
B_UreaNitrogen
B_Eosinophils
B_INR(PT)
B_MCV
B_PlateletCount
B_PT
B_RDW
B_RedBloodCells
B_WhiteBloodCells
C_LactateDehydrogenase
C_Protein
C_Eosinophils
C_Polys
C_RBC
age
resprate
acuity
Pearson Correlation
−1.0 −0.5 0.0 0.5 1.0
B
Chief Complaint (Survivors)
headache
status
weakness
pain
mental
altered
fever
transfer
seizure
confusion
Words
0 200 400 600 800 1000 1200
Word Count
C
Chief Complaint (Non-Survivors)
status
mental
altered
transfer
pain
confusion
weakness
fever
failure
headache
Words
0 10 20 30 40 50 60
Word Count
D
ICD (Survivors)
unspecified
disease
chronic
history
mention
acute
personal
kidney
specified
status
Words
0 5000 10000 15000 20000
Word Count
E
ICD (Non-Survivors)
unspecified
acute
failure
disease
chronic
kidney
specified
status
mention
classified
Words
0 500 1000 1500 2000 2500
Word Count
F
Chief Complaint
(20, 0.95)
Cumulative Frequency
Value
1.0
0.5
0.0
0 100 200 300
Token Count
G
ICD
(379, 0.95)
Cumulative Frequency
Value
1.0
0.5
0.0
0 200 400 600
Token Count

Figure 5


A
Structured data
Unstructured data
BioBERT text2vec
Embedding, Flatten
Embedding, LSTM
Embedding
Concatenate
Dense, Dropout
Dense, Dropout
Dense, Dropout
Conv1D
AveragePooling1D
Flatten, Dropout
Output
Output
Output
Multimodal_FCN
Multimodal_LSTM
Multimodal_CNN
B
Multimodal_FCN
All (AUC = 0.894)
Drop CC (AUC = 0.883)
Drop ICD (AUC = 0.863)
Drop IH (AUC = 0.910)
Drop MH (AUC = 0.895)
Drop AH (AUC = 0.904)
Drop FH (AUC = 0.896)
Final (AUC = 0.900)
C
Multimodal_LSTM
All (AUC = 0.857)
Drop CC (AUC = 0.856)
Drop ICD (AUC = 0.852)
Drop IH (AUC = 0.863)
Drop MH (AUC = 0.863)
Drop AH (AUC = 0.861)
Drop FH (AUC = 0.865)
Final (AUC = 0.857)
D
Multimodal_CNN
All (AUC = 0.932)
Drop CC (AUC = 0.918)
Drop ICD (AUC = 0.869)
Drop IH (AUC = 0.927)
Drop MH (AUC = 0.924)
Drop AH (AUC = 0.934)
Drop FH (AUC = 0.930)
Final (AUC = 0.933)
Structured Data (AUC = 0.809)
E
Independent Validation (AUC = 0.889)
TPR
FPR
0.0
0.2
0.4
0.6
0.8
1.0