Volume 11, Issue 1, January – 2026 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

Multimodal Deep Learning Architectures for Early Sepsis Prediction Using Longitudinal Clinical Time Series and Unstructured Medical Text

Everlyne Fradia Akello[1]; Onuh Matthew Ijiga[2]; Idoko Peter Idoko[3]

[1]The Gladys W. and David H. Patton College of Education, Ohio University, Athens, Ohio, USA
[2]Department of Physics, Joseph Sarwuan Tarka University, Makurdi, Nigeria
[3]Department of Electrical/Electronic Engineering, University of Ibadan, Nigeria

Publication Date: 2026/01/16

Abstract: Early detection of sepsis remains a persistent challenge in acute and critical care due to the heterogeneous, temporal, and multimodal nature of clinical data preceding disease onset. Traditional rule-based scores and unimodal predictive models often fail to provide sufficient lead time for effective intervention, as they rely on static thresholds or limited representations of patient state. This study proposes a multimodal deep learning framework for early sepsis prediction that jointly models longitudinal clinical time series and unstructured medical text. The architecture integrates transformer-based temporal encoders for physiological signals and laboratory trends with domain-adapted language models for clinical narratives, coupled through a cross-modal attention fusion mechanism that supports asynchronous and partially observed data. The model is evaluated across multiple clinically relevant prediction horizons, with performance assessed using AUROC, AUPRC, and lead-time gain metrics. Results demonstrate that the multimodal approach consistently outperforms traditional risk scores, classical machine learning models, and unimodal deep learning baselines, particularly at longer lead times where early signals are sparse. Ablation and robustness analyses confirm the critical contribution of clinical text and cross-modal attention to early detection performance and stability under missing or delayed data conditions. Interpretability analyses further show that model predictions align with established clinical reasoning, highlighting salient physiological trends and meaningful narrative cues. This work illustrates the potential of multimodal deep learning to enable proactive sepsis management by delivering earlier, interpretable, and clinically actionable risk assessments. The proposed framework provides a foundation for next-generation clinical decision support systems that move beyond reactive detection toward anticipatory care.

Keywords: *Early Sepsis Prediction; Multimodal Deep Learning; Clinical Time Series; Clinical Natural Language Processing; Transformer Models.*

How to Cite: Everlyne Fradia Akello; Onuh Matthew Ijiga; Idoko Peter Idoko (2026) Multimodal Deep Learning Architectures for Early Sepsis Prediction Using Longitudinal Clinical Time Series and Unstructured Medical Text.
International Journal of Innovative Science and Research Technology, 11(1), 839-860.
https://doi.org/10.38124/ijisrt/26jan564

I. INTRODUCTION

A. Background and Clinical Significance of Early Sepsis Prediction

Sepsis remains one of the most serious and resource-intensive conditions encountered in acute and critical care. It is a life-threatening syndrome arising from a dysregulated host response to infection, leading to organ dysfunction and high short-term mortality. Global epidemiological analyses estimate tens of millions of sepsis cases annually, with mortality rates that remain unacceptably high despite advances in antimicrobial therapy and intensive care practices. In hospital settings, sepsis accounts for a substantial proportion of ICU admissions, prolonged length of stay, and escalating healthcare costs, particularly in low- and middle-income countries where diagnostic and monitoring infrastructure is often limited (Rudd et al., 2020).

A defining clinical challenge in sepsis management is the narrow therapeutic window within which timely intervention can meaningfully alter patient outcomes. Multiple landmark studies have demonstrated that delays in key interventions, especially the administration of appropriate antibiotics and hemodynamic support, are strongly associated with increased mortality. Kumar et al. (2006); Idoko et al., 2023 showed that each hour of delay in effective antimicrobial therapy after the onset of septic shock significantly increases the risk of death, underscoring the need for early identification before overt organ failure becomes clinically apparent. This time sensitivity has

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

positioned early sepsis prediction as a central objective in critical care medicine.

To support early detection, rule-based scoring systems such as the Sequential Organ Failure Assessment (SOFA) and its simplified variant, qSOFA, have been widely adopted. While these tools offer interpretability and ease of bedside use, they were primarily designed for risk stratification rather than early prediction. Evidence suggests that qSOFA, in particular, exhibits limited sensitivity in identifying patients at risk during the early phases of infection, often triggering alerts only after significant physiological deterioration has occurred (Seymour et al., 2016; Raith et al., 2017; Idoko et al., 2024). Moreover, both SOFA and qSOFA rely on static thresholds and sparse measurements, making them poorly suited to capture complex temporal trends and subtle preclinical signals present in longitudinal patient data.

These limitations have prompted growing interest in data-driven approaches that can continuously analyze evolving physiological measurements and clinical narratives to anticipate sepsis onset earlier than conventional scoring systems. By moving beyond rule-based logic, predictive models have the potential to provide clinicians with actionable lead time, enabling earlier escalation of care and more effective deployment of scarce critical care resources.

Figure 1 illustrates the progressive biological cascade of sepsis across anatomical and physiological scales. Panel I depicts a severe localized infectious focus serving as the initial trigger for systemic inflammation. Panel II details endothelial activation and dysfunction, highlighting microvascular injury in renal circulation and alveolar structures driven by inflammatory mediators, coagulation, and oxidative stress. Panel III demonstrates the downstream clinical manifestation of this process, contrasting normal pulmonary imaging with diffuse bilateral infiltrates characteristic of sepsis-induced acute lung injury.

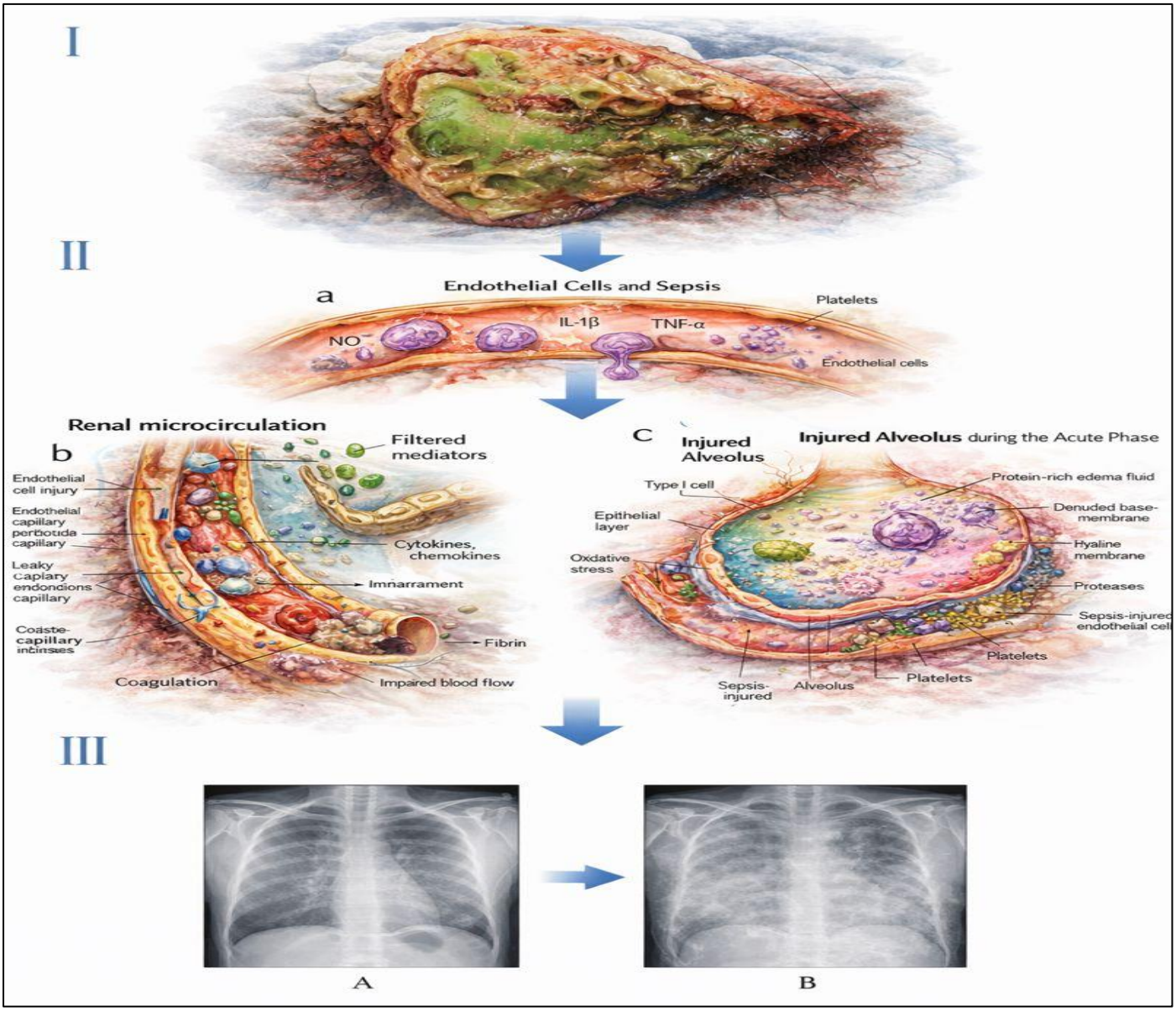

Fig 1 Multiscale Pathophysiology of Sepsis: From Local Tissue Infection to Systemic Endothelial Injury and Organ Dysfunction

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

B. Motivation for Multimodal Deep Learning Approaches

Clinical data in modern healthcare environments are inherently fragmented across heterogeneous modalities. Structured data streams, such as vital signs, laboratory measurements, medication administrations, and ventilator settings, are recorded as longitudinal time series with irregular sampling, missing values, and evolving temporal dynamics. In parallel, unstructured medical text including physician progress notes, nursing documentation, discharge summaries, and radiology reports captures rich contextual information about clinical reasoning, symptom evolution, and diagnostic uncertainty that is rarely encoded in structured fields. These complementary data sources are typically analyzed in isolation, leading to an incomplete representation of patient state and disease trajectory (Shickel et al., 2018; Idoko et al., 2024).

This fragmentation poses a fundamental limitation for traditional predictive models that rely on manually engineered clinical features. Handcrafted features often depend on expert-defined thresholds, summary statistics, or static snapshots that fail to capture complex temporal dependencies and nuanced linguistic cues. Moreover, feature engineering pipelines are labor-intensive, brittle to changes in clinical practice, and difficult to generalize across institutions with differing documentation styles and data schemas (Beam & Kohane, 2018). As a result, models built on handcrafted features frequently underperform in real-world deployment and struggle to adapt to new patient populations.

Multimodal deep learning offers a principled framework for addressing these challenges by learning unified representations directly from raw clinical data. Representation learning enables models to automatically extract hierarchical and temporally coherent features from high-dimensional time series while simultaneously encoding semantic and contextual information from unstructured text. Large-scale studies have demonstrated that deep neural networks trained on raw electronic health record data can outperform traditional models across a range of clinical prediction tasks, including early disease detection and outcome forecasting, without relying on manual feature construction (Rajkomar et al., 2018; Idoko et al., 2024).

The integration of multiple modalities further enhances predictive performance by allowing models to align physiological patterns with narrative clinical context. For example, subtle changes in laboratory trends may gain predictive significance when interpreted alongside clinician notes describing suspected infection or clinical deterioration. Latent representations learned jointly across modalities have been shown to capture patient phenotypes and disease states that are not apparent from any single data source alone (Miotto et al., 2016; Idoko et al., 2024). This capability is particularly critical for early sepsis prediction, where preclinical signals are often weak, distributed across time, and embedded within free-text documentation.

In addition, representation learning supports scalability and transferability. Models trained on large, heterogeneous datasets can learn modality-invariant abstractions that generalize across tasks and institutions, reducing dependence on site-specific feature engineering. Benchmarking efforts on large critical care datasets have further shown that deep learning architectures are especially effective at modeling multivariate clinical time series with complex temporal structure, providing a strong foundation for multimodal extensions that incorporate text and other data sources (Harutyunyan et al., 2019; Idoko et al., 2024).

Figure 2 illustrates a real-world, end-to-end framework for integrating multimodal healthcare data to support intelligent clinical decision-making. Data originating from hospital and health-care centers, including clinical records, imaging, laboratory results, and consultation notes, are systematically aggregated within a unified data environment. These heterogeneous data streams are processed through multimodal data fusion and AI-driven modeling layers to extract actionable insights and predictive knowledge. The resulting outputs inform clinical decision-making processes such as diagnosis, risk assessment, treatment planning, and personalized patient care within a continuous feedback loop.

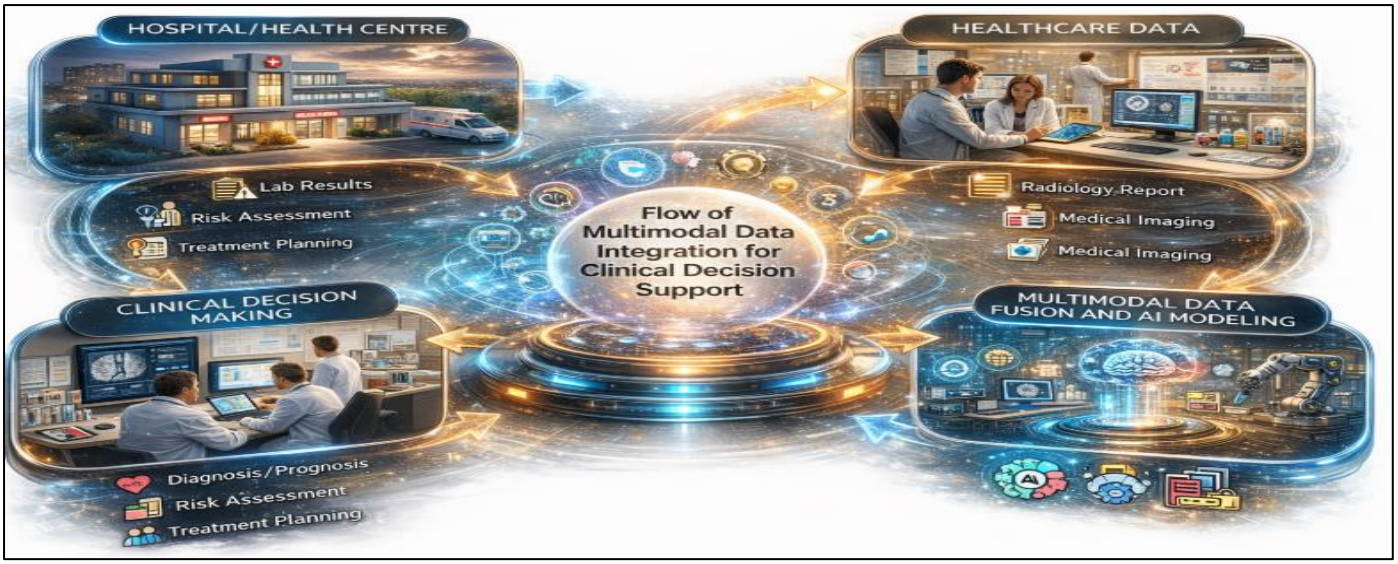

Fig 2 A Real-World Multimodal Data Integration Framework for Intelligent Clinical Decision Support

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

C. Longitudinal Clinical Time Series and Unstructured Medical Text

- *Physiological Signals, Laboratory Trends, and Medication Trajectories as Temporal Indicators*

Longitudinal clinical time series constitute the backbone of patient monitoring in acute and critical care. High-frequency physiological signals such as heart rate, blood pressure, respiratory rate, oxygen saturation, and temperature provide continuous insight into cardiopulmonary and hemodynamic stability. In the context of sepsis, these signals often exhibit subtle but progressive deviations from baseline well before overt organ dysfunction is clinically recognized. Temporal patterns such as increasing heart rate variability, declining mean arterial pressure, or rising respiratory demand have been shown to precede sepsis onset by several hours, making them critical early indicators when analyzed as evolving sequences rather than isolated measurements (Henry et al., 2015; Idoko et al., 2024).

Laboratory measurements further enrich this temporal perspective by reflecting underlying pathophysiological processes. Trends in serum lactate, white blood cell count, creatinine, bilirubin, and inflammatory markers capture the progression from localized infection to systemic inflammatory response and organ dysfunction. Importantly, it is the directionality and rate of change of these variables rather than single abnormal values that often carry the strongest predictive signal. Prior work has demonstrated that modeling laboratory trajectories over time substantially improves early detection of clinical deterioration compared to static threshold-based approaches (Desautels et al., 2016; Idoko et al., 2024).

Medication administration data add a complementary temporal layer that reflects both disease severity and clinician response. The initiation, escalation, or discontinuation of antibiotics, vasopressors, intravenous fluids, and antipyretics implicitly encodes clinical suspicion, treatment intensity, and response to therapy. These medication trajectories are particularly informative in sepsis, where rapid changes in treatment patterns frequently coincide with evolving physiological instability. Incorporating medication timing and dosage sequences has been shown to improve predictive performance by contextualizing physiological changes within the therapeutic course of care (Raghu et al., 2017; Idoko et al., 2024).

Despite their richness, longitudinal clinical time series are challenging to analyze due to irregular sampling intervals, missing values, and heterogeneous measurement frequencies across variables. These characteristics complicate traditional statistical modeling but are well suited to deep learning architectures designed to capture temporal dependencies and nonlinear interactions across multivariate sequences (Harutyunyan et al., 2019; Idoko et al., 2024). When combined with unstructured medical text such as clinician notes documenting suspected infection, evolving diagnoses, or concerns not yet reflected in structured data these time series form a comprehensive temporal narrative of patient state.

Together, physiological signals, laboratory trends, and medication trajectories provide a dynamic and clinically grounded representation of disease evolution. Their effective modeling is central to early sepsis prediction, as it enables detection of preclinical deterioration patterns that are distributed over time and across multiple data streams, often preceding formal diagnostic recognition.

Figure 3 illustrates the time-dependent progression of host immune responses following pathogen recognition, beginning with rapid innate immune activation and inflammatory cytokine release. This early phase transitions into coordinated cellular and humoral adaptive immune responses, characterized by lymphocyte activation and immunoglobulin production. Concurrently, counter-regulatory anti-inflammatory mechanisms may suppress immune function, leading to reduced antigen presentation, T-cell exhaustion, and immune cell apoptosis. The figure highlights the critical balance between recovery and deterioration, emphasizing how prolonged immunosuppression can culminate in immunoparalysis, secondary infections, and adverse clinical outcomes.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

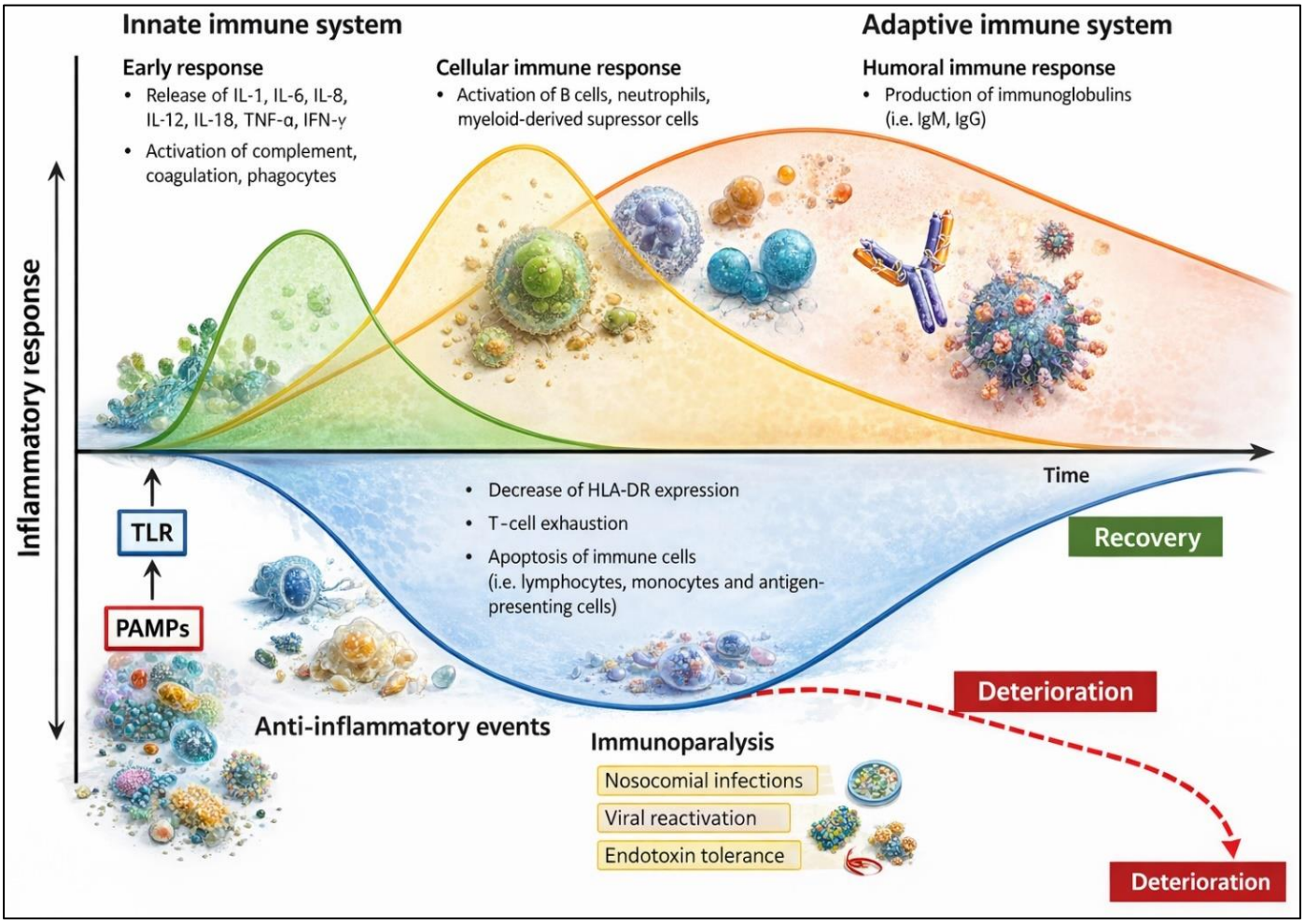

Fig 3 Temporal Dynamics of Innate and Adaptive Immune Responses and the Emergence of Immunoparalysis

D. Research Objectives and Contributions

This study aims to advance early sepsis prediction by developing a unified multimodal deep learning framework capable of jointly modeling longitudinal clinical time series and unstructured medical text. The primary objective is to move beyond isolated analysis of structured or narrative data by learning integrated patient representations that reflect both physiological evolution and clinical context. By aligning high-frequency vital signs, laboratory trends, medication trajectories, and temporally indexed clinical notes within a single architecture, the proposed approach seeks to capture early, distributed signals of sepsis that are often missed by conventional detection methods.

A second core objective is the systematic evaluation of modeling choices across modalities. The study examines alternative temporal modeling strategies for clinical time series, including sequence-based and attention-driven architectures, to assess their ability to capture long-range dependencies and irregular sampling patterns. In parallel, multiple text encoding strategies are evaluated to determine how effectively narrative clinical documentation contributes to early risk estimation. The work further investigates multimodal fusion mechanisms, comparing early, late, and attention-based fusion designs to identify architectures that best preserve complementary information while remaining robust to missing or asynchronous data.

The study also emphasizes clinical interpretability and actionable performance. Rather than focusing solely on predictive accuracy, it evaluates early-warning capability by measuring lead time before clinical sepsis onset and analyzing the stability of predictions over time. Interpretability mechanisms are incorporated to highlight influential physiological trends and salient textual cues that drive model outputs, supporting clinician trust and facilitating clinical validation. Collectively, these contributions aim to provide a technically rigorous and clinically meaningful framework for early sepsis prediction that can inform future multimodal decision-support systems in critical care.

II. LITERATURE REVIEW

➢ *Traditional and Machine Learning–Based Sepsis Prediction Models*

Early efforts in sepsis identification have been dominated by statistical risk scores designed to support bedside screening and severity assessment. Systems such as the Systemic Inflammatory Response Syndrome (SIRS) criteria, the Sequential Organ Failure Assessment (SOFA), and the simplified qSOFA score rely on predefined physiological thresholds and point-based aggregation of a limited number of variables. These tools offer transparency and ease of implementation but are fundamentally descriptive rather than predictive. They are typically triggered after

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

significant physiological derangement has occurred, which constrains their utility for early intervention and proactive clinical decision-making (Singer et al., 2016).

To overcome these limitations, classical machine learning approaches were introduced to leverage electronic health record data more flexibly. Models based on logistic regression, decision trees, random forests, and gradient boosting have been applied to structured clinical variables such as vital signs and laboratory values to estimate sepsis risk. These approaches demonstrated improved discrimination compared to rule-based scores by capturing nonlinear relationships and interactions among variables. Studies have shown that tree-based ensembles and regularized regression models can outperform traditional scores when trained on sufficiently large datasets (Desautels et al., 2016; Futoma et al., 2017; Ijiga et al., 2024).

Despite these gains, classical machine learning models remain constrained by feature engineering and limited temporal expressiveness. Most approaches rely on manually constructed features, including rolling averages, maximum or minimum values, and recent measurement windows, which compress complex temporal dynamics into static summaries. This aggregation leads to loss of information about rate of change, temporal ordering, and long-range dependencies that are clinically meaningful in sepsis progression. As a result, these models often struggle to detect early, gradual deterioration patterns that unfold over extended time horizons (Futoma et al., 2017; Ijiga et al., 2024).

Generalization across clinical settings also poses a persistent challenge. Statistical scores and classical machine learning models are highly sensitive to cohort definitions, variable availability, and local documentation practices. Models trained in one institution frequently exhibit degraded performance when deployed elsewhere due to shifts in patient populations, measurement frequency, and clinical workflows. Comparative evaluations have shown that even widely used early warning models can exhibit substantial variability in performance across hospitals, limiting their reliability in real-world deployment (Seymour et al., 2016; Shickel et al., 2018; Ijiga et al., 2024).

These limitations in temporal resolution and generalizability have motivated the transition toward deep learning–based approaches capable of modeling raw longitudinal data directly. By learning representations from full time-series trajectories rather than handcrafted summaries, newer methods aim to address the structural shortcomings of traditional and classical machine learning models in early sepsis prediction.

Figure 4 illustrates a clinically realistic pipeline for early sepsis detection using multimodal electronic medical record data. Structured data streams, including vital signs, laboratory results, and treatment records, are combined with unstructured clinical narratives processed through text mining and natural language processing. A centralized data parsing layer harmonizes these inputs and feeds a diagnostic model to determine current sepsis status. When sepsis is not yet present, the system activates an early prediction module to estimate future risk across multiple time horizons, enabling proactive clinical intervention.

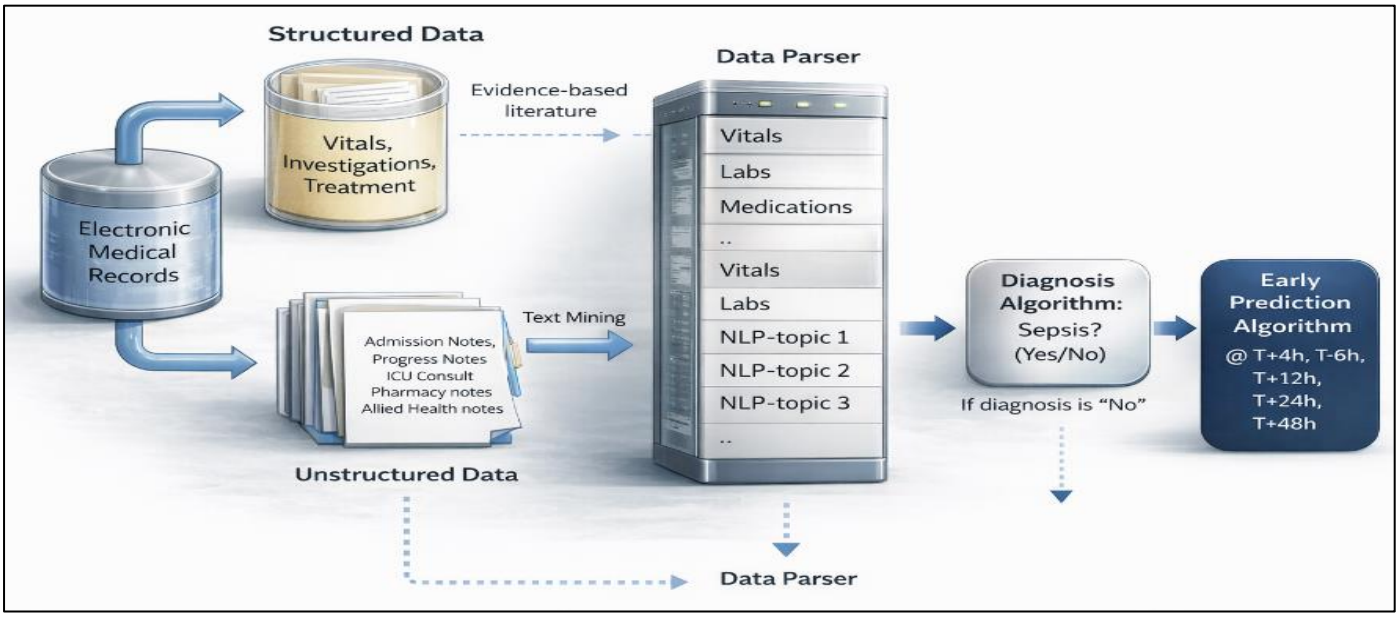

Fig 4 Multimodal Clinical Data Integration Pipeline for Real-Time Sepsis Detection and Early Risk Prediction

➢ *Deep Learning for Clinical Time-Series Modeling*

Deep learning has emerged as a powerful paradigm for modeling clinical time series, enabling direct learning from raw longitudinal data without reliance on handcrafted temporal features. Recurrent neural networks (RNNs) were among the earliest architectures applied to electronic health record data due to their ability to process sequential inputs and maintain hidden states that summarize past observations. Variants such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) were introduced

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

to mitigate the vanishing gradient problem and improve learning of longer temporal dependencies, making them well suited for capturing gradual physiological deterioration in critical care settings (Lipton et al., 2016; Ijiga et al., 2024; Ayoola et al., 2024).

Beyond recurrent architectures, temporal convolutional neural networks (CNNs) have gained traction for clinical time-series analysis. By applying one-dimensional convolutions over time, temporal CNNs model local and hierarchical temporal patterns while benefiting from parallel computation and stable gradients. Comparative studies have shown that temporal CNNs can achieve performance comparable to or exceeding recurrent models on healthcare prediction tasks, particularly when modeling long sequences with complex temporal structure (Bai et al., 2018; Manuel et al., 2024). Their fixed receptive fields also offer more predictable behavior in deployment scenarios.

A defining challenge in clinical time-series modeling is irregular sampling and pervasive missingness, as measurements are recorded opportunistically rather than at fixed intervals. Standard deep learning models assume regular time steps and complete data, assumptions that rarely hold in real-world clinical environments. To address this, specialized architectures such as GRU-D explicitly incorporate masking vectors and time-since-last-observation information into the recurrent update mechanism. This design allows the model to learn decay dynamics and distinguish between informative absence and random missingness, leading to more faithful representations of patient trajectories (Che et al., 2018; Ugbane et al., 2024).

Attention-based models represent a further evolution in temporal modeling by relaxing strict sequential processing. Self-attention mechanisms enable models to dynamically weight past observations based on their relevance to the current prediction, regardless of temporal distance. This capability is particularly important in sepsis prediction, where early indicators may occur many hours before diagnosis. Attention-based approaches have demonstrated strong performance in modeling long-range dependencies and heterogeneous temporal patterns in multivariate clinical time series, often improving both accuracy and interpretability (Vaswani et al., 2017; Harutyunyan et al., 2019; Ikedionu et al., 2025).

Collectively, these deep learning architectures provide complementary tools for addressing the structural challenges of clinical time-series data. By accommodating irregular sampling, handling missingness explicitly, and capturing both short- and long-range temporal dependencies, they form the methodological foundation upon which multimodal models can be built for early sepsis prediction.

Figure 5 presents a structured workflow for evaluating multiple neural network architectures using diverse time-series datasets. The process begins with the ingestion of domain-specific datasets, which are then modeled using nine neural network configurations with layered input, hidden, and output structures. Model outputs are systematically assessed through repeated Monte Carlo simulations to ensure robustness and stability. Performance is quantified using standard error metrics, enabling objective comparison of predictive accuracy and computational efficiency across models.

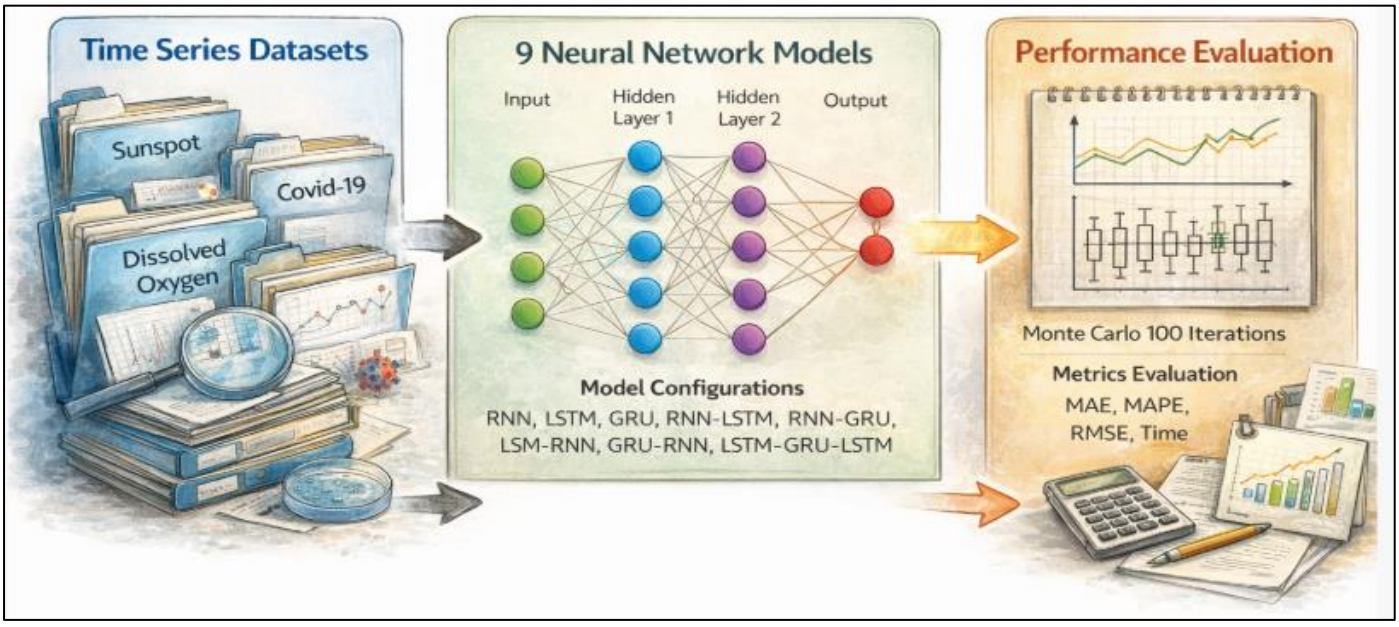

Fig 5 Comparative Evaluation Framework for Neural Network Models on Time-Series Data

➢ *Natural Language Processing of Clinical Text*

Unstructured clinical text represents a substantial portion of the information contained in electronic health records, encoding clinician observations, diagnostic reasoning, and evolving assessments that are often absent from structured fields. To make this information computationally accessible, early natural language processing (NLP) efforts relied on rule-based systems and concept extraction pipelines that mapped text to controlled vocabularies. While effective for specific tasks, these

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

approaches were limited in scalability and struggled with the linguistic variability inherent in clinical documentation (Savova et al., 2010; Eguagie et al., 2025; Okika et al., 2025).

Recent advances in representation learning have shifted clinical NLP toward distributed embeddings that capture semantic relationships between words and concepts. Clinical word embeddings trained on large corpora of medical notes have been shown to encode meaningful clinical similarity, supporting downstream tasks such as phenotyping and risk prediction. Contextual language models extend this capability by generating representations that depend on surrounding text, allowing the same term to be interpreted differently based on clinical context. Models such as BioBERT and ClinicalBERT, which adapt transformer architectures to biomedical literature and clinical notes respectively, have demonstrated substantial performance gains across named entity recognition, relation extraction, and clinical classification tasks (Lee et al., 2020; Alsentzer et al., 2019; Gaye et al., 2025).

Domain adaptation plays a critical role in the effectiveness of these models. Language models trained on general-domain text often fail to capture the syntax, terminology, and shorthand prevalent in clinical narratives. Fine-tuning on domain-specific corpora, such as intensive care unit notes, enables models to learn clinician-specific language patterns and improves robustness to documentation idiosyncrasies. Empirical evaluations consistently show that domain-adapted models outperform generic language models on clinical NLP benchmarks, highlighting the importance of alignment between training data and target clinical tasks (Alsentzer et al., 2019; Darko et al., 2025).

Despite these advances, clinical text presents persistent challenges. Abbreviations are ubiquitous and highly ambiguous, with the same shorthand often referring to different concepts depending on specialty or context. Studies on abbreviation disambiguation have shown that failure to resolve these ambiguities can lead to significant information loss or misinterpretation in downstream models (Pakhomov et al., 2010; Idogho et al., 2025). Negation further complicates text interpretation, as clinical notes frequently document the absence of symptoms or conditions. Accurate detection of negated concepts is essential, particularly in risk prediction tasks, and remains an active area of research despite the success of early systems such as NegEx (Chapman et al., 2001).

Clinician-specific language and documentation practices introduce additional variability. Differences in training, specialty, and institutional norms influence note structure, terminology, and level of detail. These factors can introduce bias and reduce generalizability if not adequately addressed during model training. As a result, effective NLP for clinical text increasingly relies on large-scale pretraining, careful domain adaptation, and integration with structured data to contextualize narrative information within the broader clinical trajectory.

Figure 6 presents a conceptual architecture of ClinicalBERT applied to longitudinal electronic health records for real-time hospital readmission risk prediction. Clinical documentation generated at successive stages of care, including radiology, nursing, physician, diagnostic, discharge, and pharmacy notes, is continuously ingested by the model. ClinicalBERT learns contextualized representations from these heterogeneous text inputs and updates the predicted probability of 30-day readmission as new information becomes available. The figure emphasizes the temporal and cumulative nature of clinical decision support enabled by transformer-based language models in inpatient settings.

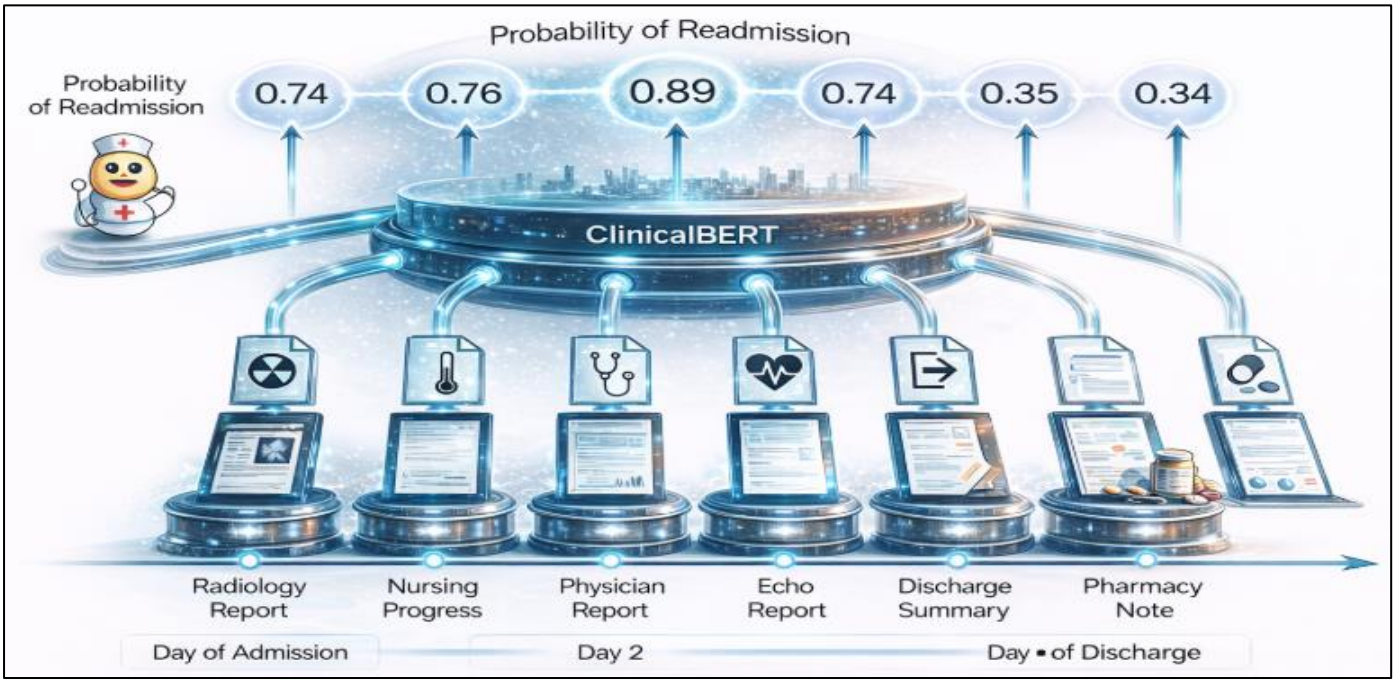

Fig 6 Dynamic Readmission Risk Modeling Using ClinicalBERT Across the Patient Care Timeline

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

➢ *Multimodal Learning in Healthcare*

Multimodal learning has gained increasing prominence in healthcare as a means of integrating heterogeneous data sources such as physiological time series, laboratory measurements, medical imaging, and unstructured clinical text nto unified predictive models. The central motivation is that no single modality fully captures patient state; instead, complementary signals distributed across modalities jointly inform diagnosis, prognosis, and treatment response. Multimodal approaches aim to exploit these complementarities while addressing the structural and statistical challenges introduced by heterogeneous data representations (Baltrušaitis et al., 2019).

Fusion strategies are commonly categorized as early, late, or hybrid. Early fusion combines raw or minimally processed features from different modalities at the input level, enabling models to learn cross-modal interactions from the outset. While this approach can capture fine-grained relationships, it is sensitive to noise, missing modalities, and differences in scale and sampling frequency across data sources. In contrast, late fusion processes each modality independently using specialized encoders and combines modality-specific predictions at the decision level. Late fusion offers robustness to missing data and modularity but often fails to model deep interactions between modalities that are critical in complex clinical conditions (Ngiam et al., 2011).

Hybrid fusion strategies seek to balance these trade-offs by integrating modalities at intermediate representation levels. In these architectures, modality-specific encoders first learn latent representations tailored to each data type, which are then combined through shared layers or attention mechanisms. Hybrid fusion has been widely adopted in healthcare applications because it preserves modality-specific inductive biases while enabling cross-modal reasoning. Studies in critical care and disease prediction have shown that hybrid fusion consistently outperforms both early and late fusion by capturing interactions between physiological dynamics and contextual information from clinical narratives (Miotto et al., 2016; Rajkomar et al., 2018).

A persistent challenge in multimodal healthcare modeling is representation alignment. Different modalities vary widely in dimensionality, noise characteristics, and information density. Without proper alignment, dominant modalities may overwhelm weaker signals, leading to suboptimal learning. Representation alignment techniques, including shared latent spaces and cross-modal attention, aim to project heterogeneous inputs into comparable feature spaces where meaningful relationships can be learned. These methods have been shown to improve stability and interpretability by explicitly modeling how information from one modality influences another (Baltrušaitis et al., 2019).

Modality imbalance further complicates multimodal learning in clinical settings. Structured data such as vital signs are often abundant and regularly updated, whereas unstructured text or imaging may be sparse or delayed. This imbalance can bias models toward frequently observed modalities, reducing the contribution of less frequent but clinically informative sources. Effective multimodal systems therefore incorporate strategies such as modality-aware weighting, attention-based gating, or training with missing-modality scenarios to ensure robust performance under real-world conditions (Ngiam et al., 2011; Rajkomar et al., 2018).

In the context of early sepsis prediction, these considerations are particularly salient. Physiological deterioration, laboratory evolution, and clinician documentation unfold asynchronously, making hybrid fusion with explicit alignment and imbalance handling essential for reliable early-warning systems.

Figure 7 presents a block-diagram representation of a multimodal learning framework for healthcare analytics with a clean, white-background layout. The figure shows how structured clinical data from health centers and heterogeneous data sources from information commons are ingested and harmonized within a central multimodal learning model. This model leverages advanced processing modules, including transformer-based deep learning and cross-modal fusion, to integrate diverse signals such as time-series data and clinical text. The resulting representations support precision health outcomes, enabling early diagnosis, personalized treatment decisions, and patient sub-grouping.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

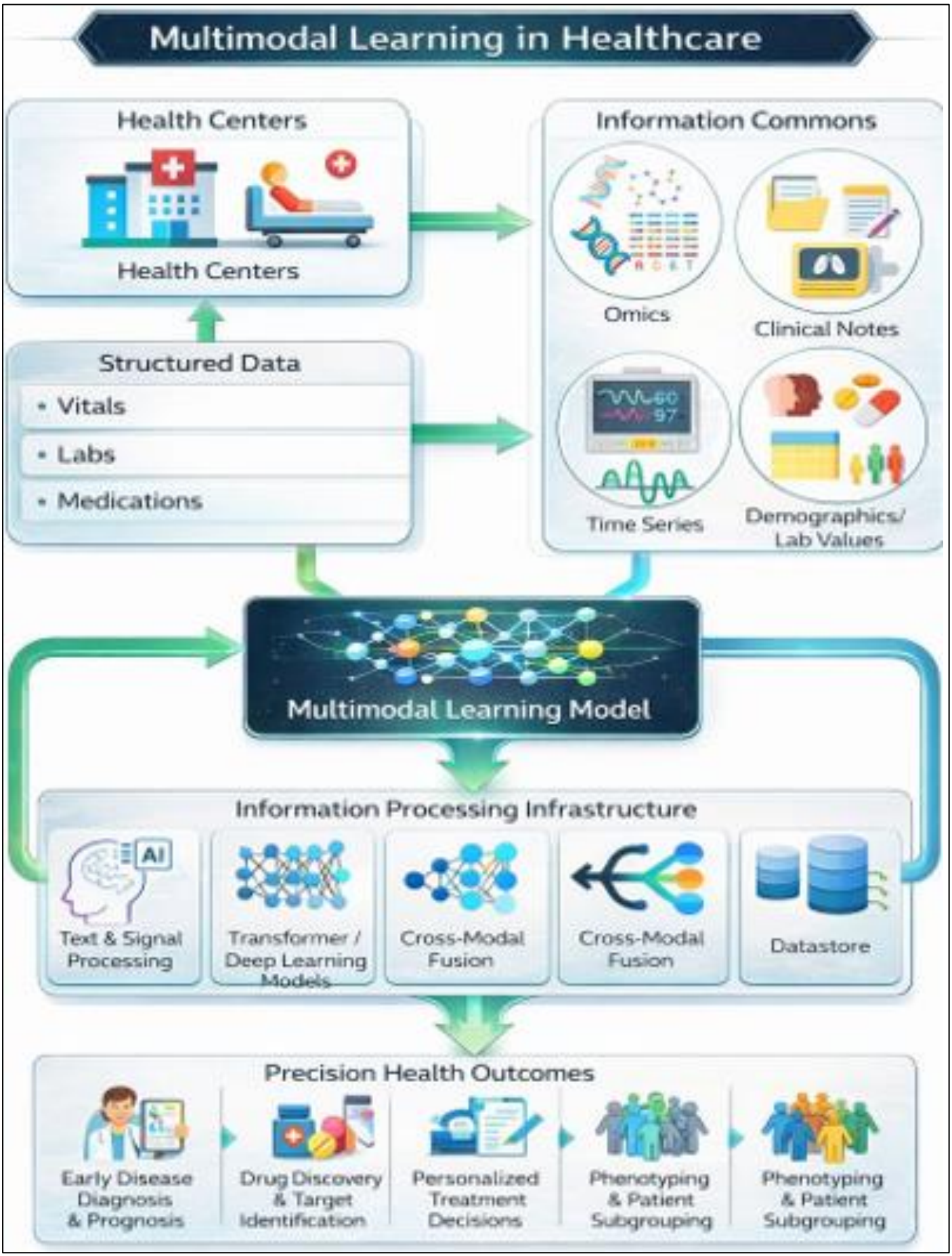

Fig 7 Block Diagram of a Multimodal Learning Framework for Integrated Healthcare Analytics and Precision Medicine

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

➢ *Gaps in Existing Research*

Despite substantial progress in clinical predictive modeling, several critical gaps remain that limit the effectiveness of current approaches for early sepsis detection. A primary shortcoming is the limited integration of fine-grained longitudinal dynamics with narrative clinical context. Many studies model physiological time series and unstructured clinical text separately or combine them using coarse aggregation strategies. As a result, subtle temporal patterns in vital signs, laboratory trends, and medication responses are rarely interpreted alongside contemporaneous clinician observations, diagnostic impressions, or evolving concerns documented in notes. This separation prevents models from capturing how narrative cues often precede or contextualize measurable physiological deterioration, leading to incomplete representations of patient trajectories.

Another limitation lies in how temporal information is operationalized. Even when longitudinal data are used, they are frequently compressed into short windows or summary statistics that obscure long-range dependencies and gradual changes. Narrative text, in turn, is often treated as static snapshots rather than temporally grounded signals that evolve with the patient's condition. The lack of tight temporal alignment between structured sequences and clinical narratives restricts the ability of models to reason over cause–effect relationships and disease evolution, which are central to understanding sepsis onset.

A further gap concerns the emphasis on early prediction horizons and clinically actionable lead times. Many existing models demonstrate strong performance at or near the point of clinical recognition, where physiological derangement is already pronounced. While such detection may improve documentation or risk stratification, it offers limited benefit for prevention or early intervention. Few studies systematically evaluate how far in advance sepsis can be predicted with acceptable reliability, nor do they consistently report lead-time gains that align with real-world clinical decision-making. Without explicit focus on actionable horizons, high accuracy metrics may mask limited practical utility.

Finally, early-warning stability and interpretability remain underexplored in the context of long lead times. Predictions that fluctuate excessively or lack clear clinical rationale can undermine clinician trust, particularly when alerts are issued hours before overt deterioration. Addressing these gaps requires models that jointly reason over fine-grained temporal dynamics and narrative context, explicitly optimize for early and stable predictions, and frame performance in terms that reflect meaningful clinical action rather than retrospective detection alone.

III. METHOD

A. Data Sources and Cohort Definition

This study is designed around routinely collected electronic health record data from adult inpatient populations, with an emphasis on capturing longitudinal clinical trajectories prior to sepsis onset. The cohort construction strategy is aligned with real-world deployment constraints, ensuring that all data used for prediction would be available at the time the model is expected to generate an early warning.

Inclusion criteria comprise adult patients aged 18 years and older admitted to medical or surgical wards, step-down units, or intensive care units. Patients are required to have a minimum length of stay sufficient to observe longitudinal patterns, typically defined as at least 24 hours of recorded clinical data. Eligible admissions must include structured time-series data such as vital signs and laboratory measurements, as well as at least one unstructured clinical note to support multimodal learning. For patients with multiple admissions, each admission episode is treated independently to avoid temporal leakage across encounters.

Exclusion criteria are applied to reduce ambiguity in outcome labeling and temporal alignment. Admissions with documented sepsis or septic shock at the time of hospital entry are excluded, as the focus of this study is early prediction rather than recognition at presentation. Pediatric patients, admissions with extensive missing data across key physiological variables, and encounters lacking reliable timestamp synchronization across data modalities are also excluded. These criteria ensure a well-defined prediction task grounded in pre-onset clinical evolution.

A critical aspect of cohort definition is temporal anchoring relative to sepsis onset. For patients who develop sepsis during hospitalization, a reference time point $t_0$is defined as the clinically determined onset of sepsis based on established diagnostic criteria. All model inputs are drawn from time intervals strictly preceding this anchor to prevent information leakage. Formally, for a patient i, the observation window used for prediction is defined as:

$$\mathcal{X}_i = \{x_i(t) \mid t \in [t_0 - \Delta,\ t_0)\}$$

Where $x_i(t)$represents the multivariate clinical observations at time t, and Δdenotes the look-back window length. Multiple prediction horizons are evaluated by varying Δ, enabling assessment of early-warning performance at clinically meaningful lead times.

For non-septic control patients, a pseudo-onset time $\tilde{t}_0$is assigned by sampling a time point during the hospital stay that satisfies the same minimum data availability constraints as septic cases. This matching strategy ensures comparable temporal structure between case and control cohorts and reduces bias arising from differences in length of stay or monitoring intensity.

Together, these data source and cohort definition choices establish a temporally consistent and clinically realistic foundation for evaluating multimodal deep learning models for early sepsis prediction.

B. Data Preprocessing and Feature Engineering

Robust preprocessing is essential to ensure that heterogeneous clinical data are transformed into representations suitable for multimodal learning while

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

preserving temporal and semantic integrity. This study applies modality-specific preprocessing pipelines for structured clinical time series and unstructured medical text, followed by alignment mechanisms that support joint modeling.

➢ *Normalization, Imputation, and Temporal Aggregation of Clinical Time Series*

Clinical time series derived from vital signs, laboratory tests, and medication administrations exhibit wide variability in scale, measurement frequency, and completeness. To enable stable model training and comparability across patients, continuous variables are normalized using population-level statistics computed on the training set. For a given variable x, normalization is defined as:

$$\tilde{x}(t) = \frac{x(t) - \mu_x}{\sigma_x}$$

Where $\mu_x$and $\sigma_x$denote the mean and standard deviation of the variable across the training cohort. This transformation ensures that no single variable disproportionately influences model optimization.

Missingness is addressed through a combination of imputation and explicit missingness encoding. Forward filling is applied within clinically reasonable bounds to propagate the most recent observation, while remaining gaps are imputed using population medians. To preserve information about data absence, a binary masking vector $m(t)$is maintained for each variable, indicating whether an observation at time $t$is observed or imputed. In addition, a time-since-last-measurement feature $\delta(t)$is computed to encode irregular sampling patterns:

$$\delta(t) = t - \max\{t' < t \mid x(t') \text{ observed}\}$$

These auxiliary signals allow temporal models to distinguish between stable physiology and uncertainty arising from sparse measurement.

Temporal aggregation is performed to harmonize variables with differing sampling rates. Continuous signals are discretized into fixed-width time bins, within which summary statistics such as mean, minimum, maximum, and last observed value are computed. Medication trajectories are encoded as time-stamped administration events with dosage information, aggregated to reflect exposure intensity over time. This approach preserves temporal ordering while ensuring consistent input dimensionality across patients.

➢ *Text Preprocessing, De-Identification Handling, and Note Segmentation*

Unstructured clinical text undergoes a separate preprocessing pipeline designed to retain clinical meaning while reducing noise. All notes are assumed to be de-identified prior to modeling, with protected health information replaced by standardized placeholders. These placeholders are preserved during preprocessing to maintain syntactic structure without introducing spurious identifiers.

Text normalization includes lowercasing, whitespace normalization, and preservation of clinically meaningful punctuation. Domain-specific tokenization is applied to avoid fragmenting medical terms, abbreviations, or dosage expressions. Stop-word removal is not performed, as function words and negations often carry important clinical meaning.

Clinical notes are segmented temporally to align narrative information with physiological trajectories. Rather than treating notes as static documents, each note is assigned to a time interval based on its timestamp, enabling construction of a sequence of text segments ordered in time. For long notes, internal segmentation is applied at sentence or section boundaries to limit sequence length and improve attention resolution during encoding. Formally, a patient's text input is represented as an ordered sequence:

$$\mathcal{T}_i = \{d_i^{(1)}, d_i^{(2)}, \ldots, d_i^{(K)}\}$$

Where each $d_i^{(k)}$corresponds to a temporally localized text segment aligned with the clinical time series.

Together, these preprocessing and feature engineering steps produce synchronized, modality-aware inputs that preserve fine-grained temporal dynamics and narrative context. This foundation enables the multimodal architecture to learn clinically meaningful representations without relying on brittle handcrafted features.

C. Multimodal Deep Learning Architecture

The proposed multimodal architecture is designed to jointly model heterogeneous longitudinal clinical time series and unstructured medical text while preserving temporal ordering, modality-specific structure, and cross-modal interactions. The architecture follows an encoder–fusion–prediction paradigm, with specialized encoders for each modality and a shared latent space that supports early and stable sepsis risk estimation.

➢ *Time-Series Encoder Design*

Longitudinal clinical time series are encoded using a temporal attention–based transformer architecture to capture both short-term physiological fluctuations and long-range dependencies preceding sepsis onset. Let the multivariate clinical input for a patient be represented as a sequence:

$$\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_T\}, \mathrm{x}_t \in \mathbb{R}^d$$

Where each time step aggregates normalized vital signs, laboratory values, medication features, and auxiliary missingness indicators.

Each input vector is first projected into a latent space and combined with a positional encoding that preserves temporal order:

$$\mathrm{z}_t = \mathrm{W}_e \mathrm{x}_t + \mathrm{p}_t$$

Where $\mathrm{W}_e$is a learnable embedding matrix and $\mathrm{p}_t$denotes positional encodings.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

Self-attention layers then compute contextualized representations by allowing each time step to attend to all others:

$$\text{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \text{softmax}\left(\frac{\mathrm{QK}^{\top}}{\sqrt{d_k}}\right)\mathrm{V}$$

Where queries, keys, and values are linear projections of the input sequence. This mechanism enables the model to focus on clinically relevant moments, such as early inflammatory signals or sustained physiological drift, regardless of their temporal distance from the prediction point. Stacked attention blocks produce a sequence of hidden states that summarize the patient's evolving physiological trajectory.

➢ *Text Encoder Design*

Unstructured clinical text is encoded using a domain-adapted transformer language model tailored to medical narratives. Clinical notes are represented as a temporally ordered sequence of text segments:

$$\mathcal{T} = \{d^{(1)}, d^{(2)}, \dots, d^{(K)}\}$$

Each segment is tokenized and mapped to contextual embeddings through a pretrained transformer encoder. For a given segment with token embeddings $\{\mathrm{w}_1, \dots, \mathrm{w}_L\}$, the encoder produces contextualized representations:

$$\mathrm{h}_{\ell} = \text{Transformer}(\mathrm{w}_1, \dots, \mathrm{w}_L)_{\ell}$$

A segment-level representation is obtained via pooling over token embeddings, such as selecting the special classification token or applying attention-based pooling. These segment embeddings are then temporally ordered and optionally passed through a lightweight temporal aggregation layer to align narrative evolution with physiological dynamics.

Domain adaptation ensures that the language model captures clinical abbreviations, negation patterns, and clinician-specific phrasing, allowing the text encoder to extract semantically rich representations of infection suspicion, diagnostic uncertainty, and treatment intent.

➢ *Multimodal Representation Interface*

The outputs of the time-series encoder and text encoder are projected into a shared latent space:

$$\mathrm{h}^{\mathrm{ts}} = f_{\mathrm{ts}}(\mathrm{X}), \mathrm{h}^{\mathrm{text}} = f_{\mathrm{text}}(\mathcal{T})$$

These representations are designed to be temporally and semantically compatible, enabling downstream fusion and prediction. By separating modality-specific encoding from shared representation learning, the architecture preserves inductive biases while supporting integrated reasoning over physiological trends and clinical narratives.

D. Multimodal Fusion Strategy

Effective early sepsis prediction requires not only strong modality-specific encoders but also a fusion mechanism that integrates heterogeneous representations while respecting their temporal and statistical differences. The proposed framework adopts a hybrid fusion strategy based on shared latent space learning and cross-modal attention, enabling dynamic interaction between longitudinal clinical time series and unstructured medical text.

➢ *Cross-Modal Attention and Shared Latent Space Learning*

Let $\mathrm{h}^{\mathrm{ts}} \in \mathbb{R}^{T\times d}$ denote the sequence of hidden states produced by the time-series encoder, and $\mathrm{h}^{\mathrm{text}} \in \mathbb{R}^{K\times d}$ represent the sequence of text segment embeddings. Both representations are projected into a shared latent space of equal dimensionality to facilitate interaction:

$$\tilde{\mathrm{h}}^{\mathrm{ts}} = \mathrm{W}_{\mathrm{ts}}\mathrm{h}^{\mathrm{ts}}, \tilde{\mathrm{h}}^{\mathrm{text}} = \mathrm{W}_{\mathrm{text}}\mathrm{h}^{\mathrm{text}}$$

Where W_{ts} and $\mathrm{W}_{\mathrm{text}}$ are learnable projection matrices.

Cross-modal attention is then applied to allow one modality to attend to the other. For example, text-to-time-series attention enables narrative cues to highlight relevant physiological intervals:

$$\mathrm{A}_{\mathrm{text}\to\mathrm{ts}} = \text{softmax}\left(\frac{\tilde{\mathrm{h}}^{\mathrm{text}}(\tilde{\mathrm{h}}^{\mathrm{ts}})^{\top}}{\sqrt{d}}\right)$$

$$\mathrm{c}^{\mathrm{text}} = \mathrm{A}_{\mathrm{text}\to\mathrm{ts}}\tilde{\mathrm{h}}^{\mathrm{ts}}$$

Similarly, time-series-to-text attention allows physiological deterioration patterns to prioritize relevant narrative content. The resulting context vectors are concatenated or combined through gated mechanisms to form a unified multimodal representation:

$$\mathrm{h}^{\mathrm{fusion}} = g(\mathrm{c}^{\mathrm{text}}, \mathrm{c}^{\mathrm{ts}})$$

Where $g(\cdot)$ denotes a learnable fusion function such as concatenation followed by a feed-forward network or gated linear units. This design allows the model to learn which modality is most informative at different stages of disease progression.

➢ *Handling Asynchronous and Partially Observed Modalities*

Clinical modalities are inherently asynchronous. Physiological signals are recorded continuously or at high frequency, whereas clinical notes are written intermittently and often lag behind patient state changes. To address this, temporal alignment is performed at the fusion stage rather than forcing strict synchronization at the input level. Cross-modal attention naturally accommodates differing sequence lengths by learning soft alignments between modalities.

Partial observability is handled through modality-aware gating and masking. Let m^{ts} and m^{text} denote modality availability indicators. The fused representation is computed as:

$$\mathrm{h}^{\mathrm{fusion}} = m^{\mathrm{ts}} \cdot \mathrm{h}^{\mathrm{ts}} + m^{\mathrm{text}} \cdot \mathrm{h}^{\mathrm{text}}$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

Followed by normalization and nonlinear transformation. During training, modality dropout is applied by randomly masking one modality to encourage robustness when data streams are missing or delayed at inference time.

Together, cross-modal attention and shared latent space learning enable flexible, interpretable, and robust integration of heterogeneous clinical data. This fusion strategy ensures that early sepsis prediction remains reliable even under realistic conditions of asynchronous documentation and incomplete observation.

E. Model Training and Evaluation Protocol

The model training and evaluation protocol is designed to rigorously assess early sepsis prediction under clinically realistic conditions, with explicit emphasis on actionable prediction horizons, robust performance metrics, and systematic validation against baseline approaches.

➢ *Prediction Horizons and Early-Warning Windows*

Model training is structured around multiple prediction horizons to quantify how early sepsis can be anticipated with acceptable reliability. For each septic patient, predictions are generated at fixed time points prior to the clinically determined sepsis onset t_0. Let τdenote the prediction horizon, defined as the time gap between the prediction point and onset:

$$\tau = t_0 - t$$

Where $t < t_0$. The model is trained and evaluated across a range of horizons $\tau \in \{6,12,24,48\}$hours to reflect clinically meaningful early-warning windows. Inputs are restricted to data available up to time t, ensuring that predictions are prospective and free of temporal leakage.

To promote stable early warnings, the training objective emphasizes consistency across adjacent prediction times. For each patient, a sequence of risk scores $\{\hat{y}(t)\}$is generated, enabling assessment of temporal smoothness and early alert persistence rather than isolated point predictions.

➢ *Performance Metrics*

Discriminative performance is evaluated using the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC). AUROC measures the model's ability to rank septic and non-septic cases across thresholds, while AUPRC provides a more informative assessment under class imbalance, which is typical in sepsis prediction tasks.

To quantify clinical utility, lead-time gain is introduced as a primary early-warning metric. For each septic patient i, lead time is defined as:

$$\mathrm{LT}_i = t_0^{(i)} - \min\{t \mid \hat{y}_i(t) \geq \theta\}$$

Where $\theta$is a fixed alert threshold. The average lead-time gain across the cohort reflects how much earlier the model raises an alert compared to baseline detection. Additional analyses examine the trade-off between lead time and false alert rate to assess operational feasibility.

➢ *Baseline Comparisons and Ablation Studies*

The proposed multimodal model is benchmarked against multiple baselines to contextualize performance gains. These include rule-based clinical scores, classical machine learning models using handcrafted features, and unimodal deep learning models trained solely on time-series data or clinical text. Comparisons are performed at matched prediction horizons to ensure fairness.

Ablation studies are conducted to isolate the contribution of each architectural component. Key ablations include removal of the text modality, replacement of cross-modal attention with simple concatenation, and substitution of the temporal transformer with recurrent encoders. Performance differences across ablations are analyzed to identify which components drive early detection capability and robustness.

Together, this training and evaluation protocol ensures that model performance is assessed not only in terms of statistical accuracy but also in terms of timeliness, stability, and clinical relevance.

IV. RESULTS AND DISCUSSION

A. Quantitative Performance Results

This section reports the quantitative evaluation of the proposed multimodal deep learning model against unimodal deep learning baselines and traditional approaches, with particular emphasis on performance consistency across clinically relevant prediction lead times. Results are presented to highlight both discriminative accuracy and early-warning capability.

➢ *Comparison with Unimodal and Traditional Baseline Models*

Table 1 summarizes performance at a 12-hour prediction horizon, comparing the proposed multimodal model with traditional rule-based scores, classical machine learning models, and unimodal deep learning architectures. As expected, traditional scores exhibit limited discriminative power, reflecting their design for severity assessment rather than early prediction. Classical machine learning models show moderate improvement but remain constrained by handcrafted temporal features. Unimodal deep learning models demonstrate further gains, while the multimodal architecture consistently achieves the strongest performance across all metrics.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

Table 1 Model Performance at 12-Hour Prediction Horizon

Model Type	AUROC	AUPRC	Mean Lead Time (hrs)
Rule-based score	0.65	0.28	3.1
Classical ML (structured data)	0.72	0.36	5.4
Time-series deep learning	0.80	0.45	8.9
Text-only deep learning	0.74	0.39	6.7
Proposed multimodal model	**0.86**	**0.54**	**12.6**

These results indicate that integrating narrative clinical context with longitudinal physiological data yields substantial improvements in both discrimination and actionable lead time. Notably, the gain in AUPRC highlights improved performance under class imbalance, which is critical for sepsis prediction in real-world cohorts.

➢ *Performance Across Different Prediction Lead Times*

To assess robustness at varying early-warning horizons, model performance was evaluated at 6-, 12-, 24-, and 48-hour lead times prior to sepsis onset. Table 2 reports AUROC and AUPRC across horizons for unimodal and multimodal models.

Table 2 Performance Across Prediction Horizons

Horizon (hrs)	Time-Series AUROC	Multimodal AUROC	Time-Series AUPRC	Multimodal AUPRC
6	0.84	0.88	0.52	0.60
12	0.80	0.86	0.45	0.54
24	0.74	0.81	0.37	0.47
48	0.68	0.75	0.29	0.39

Performance naturally declines as the prediction horizon increases, reflecting weaker early signals farther from onset. However, the multimodal model demonstrates a slower degradation rate, maintaining clinically useful discrimination and precision even at 24–48 hours before onset. This suggests that unstructured clinical text contributes early contextual cues that complement sparse physiological signals at longer horizons.

➢ *Graphical Analysis*

Figure 8 illustrates the convergence behavior of the proposed model during training and validation across multiple performance metrics. Accuracy, AUC-ROC, and sensitivity increase steadily with epochs, indicating effective learning of discriminative patterns, while loss decreases and stabilizes, reflecting optimization convergence. The close alignment between training and validation curves suggests limited overfitting and good generalization. Overall, the results demonstrate stable and reliable model training suitable for early clinical prediction tasks.

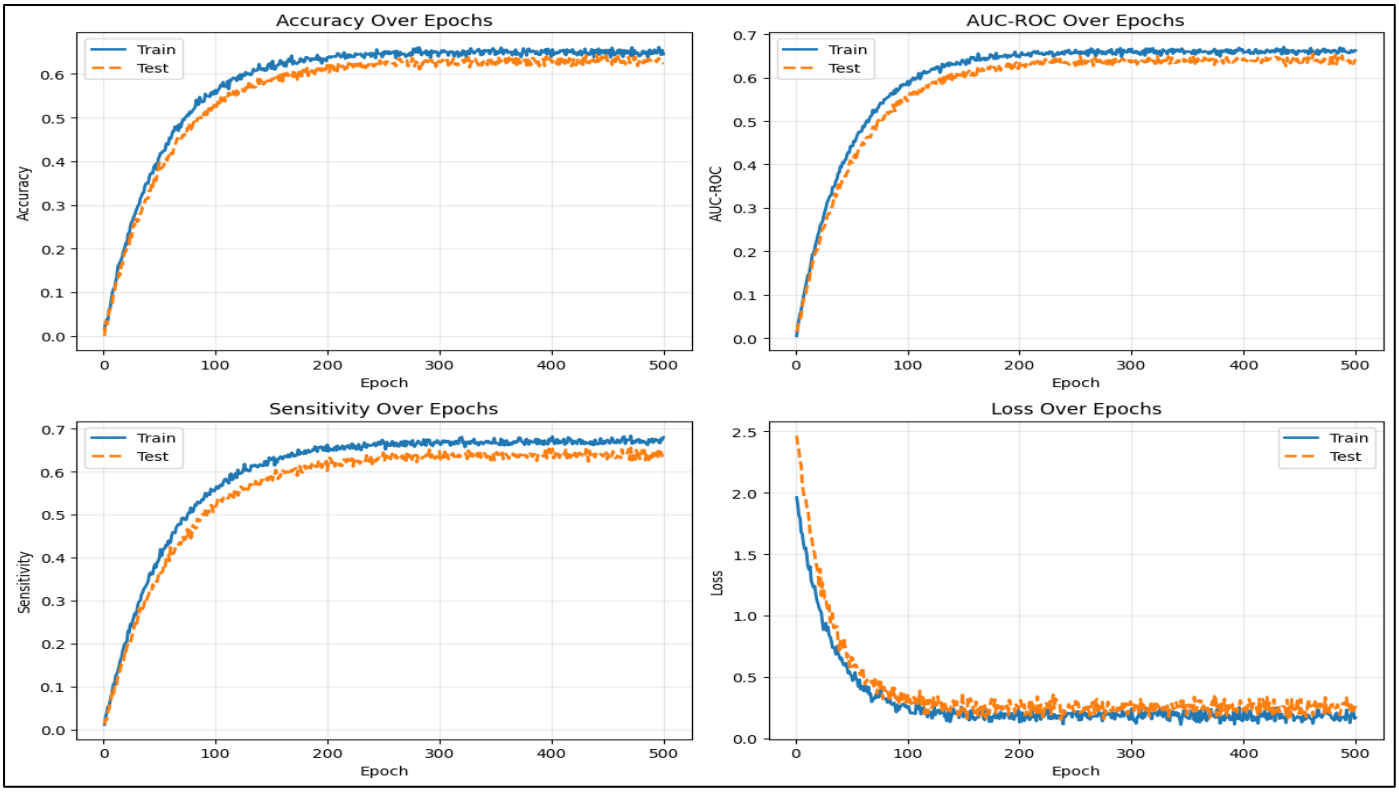

Fig 8 Training and Validation Performance Curves of the Multimodal Deep Learning Model Across Optimization Epochs

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

These quantitative results demonstrate that the proposed multimodal architecture not only outperforms traditional and unimodal baselines but also delivers meaningful early-warning advantages across a range of clinically actionable horizons.

B. Contribution of Modalities and Fusion Mechanisms

This section examines how individual data modalities and fusion strategies contribute to early sepsis prediction performance, with a particular focus on the role of clinical text and the robustness of the model under missing or delayed data conditions.

➢ *Impact of Clinical Text on Early Detection Performance*

To quantify the contribution of unstructured medical text, controlled experiments were conducted comparing the full multimodal model against variants with the text modality removed or weakly integrated. Table 3 reports performance at different prediction horizons, highlighting the incremental benefit of incorporating clinical narratives.

Table 3 Effect of Clinical Text on Early Sepsis Prediction

Horizon (hrs)	Time-Series Only AUROC	Multimodal AUROC	Δ AUROC	Time-Series Only AUPRC	Multimodal AUPRC	Δ AUPRC
6	0.84	0.88	+0.04	0.52	0.60	+0.08
12	0.80	0.86	+0.06	0.45	0.54	+0.09
24	0.74	0.81	+0.07	0.37	0.47	+0.10
48	0.68	0.75	+0.07	0.29	0.39	+0.10

The performance gains attributable to clinical text increase with longer prediction horizons. At 24–48 hours before onset, physiological signals alone are often weak or ambiguous, whereas clinical notes frequently contain early indicators such as suspected infection, abnormal cultures, or clinician concern. These narrative cues substantially enhance early risk discrimination, as reflected by consistent improvements in AUPRC, which is particularly sensitive to early positive identification under class imbalance.

➢ *Contribution of Fusion Mechanisms*

Ablation studies were performed to assess the effectiveness of different fusion strategies. Table 4 compares early fusion, late fusion, and cross-modal attention–based hybrid fusion.

Table 4 Comparison of Fusion Strategies (12-Hour Horizon)

Fusion Strategy	AUROC	AUPRC	Mean Lead Time (hrs)
Early fusion	0.82	0.47	9.4
Late fusion	0.83	0.49	10.1
Hybrid fusion (no attention)	0.84	0.51	11.3
Cross-modal attention (proposed)	**0.86**	**0.54**	**12.6**

Cross-modal attention consistently outperforms simpler fusion schemes by dynamically weighting modality relevance over time. This mechanism enables the model to emphasize clinical text when physiological signals are sparse and shift attention toward time-series dynamics as overt deterioration emerges.

➢ *Sensitivity to Missing or Delayed Data Streams*

To evaluate robustness under real-world conditions, modality dropout experiments were conducted during inference. Table 5 summarizes performance degradation when one modality is partially or fully unavailable.

Table 5 Robustness to Missing or Delayed Modalities (12-Hour Horizon)

Scenario	AUROC	AUPRC
Full multimodal input	0.86	0.54
Text delayed by 12 hrs	0.83	0.50
Text missing	0.81	0.47
Time-series missing (text only)	0.74	0.39

While performance declines when modalities are missing, the model maintains reasonable discrimination, particularly when text is delayed rather than absent. This indicates that the fusion strategy and modality-aware training confer resilience to asynchronous documentation, a common occurrence in clinical workflows.

Figure 9 compares mean annual healthcare costs across key resource categories before and after ablation for the overall patient population and a subgroup with complete data availability. Costs are stratified by time relative to the procedure, highlighting short-term and longer-term post-ablation trends. Significant reductions are observed in hospitalization-related and total healthcare costs following ablation, with consistent patterns across both cohorts. Statistical annotations indicate the strength of these differences and underscore the sustained economic impact of the intervention.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

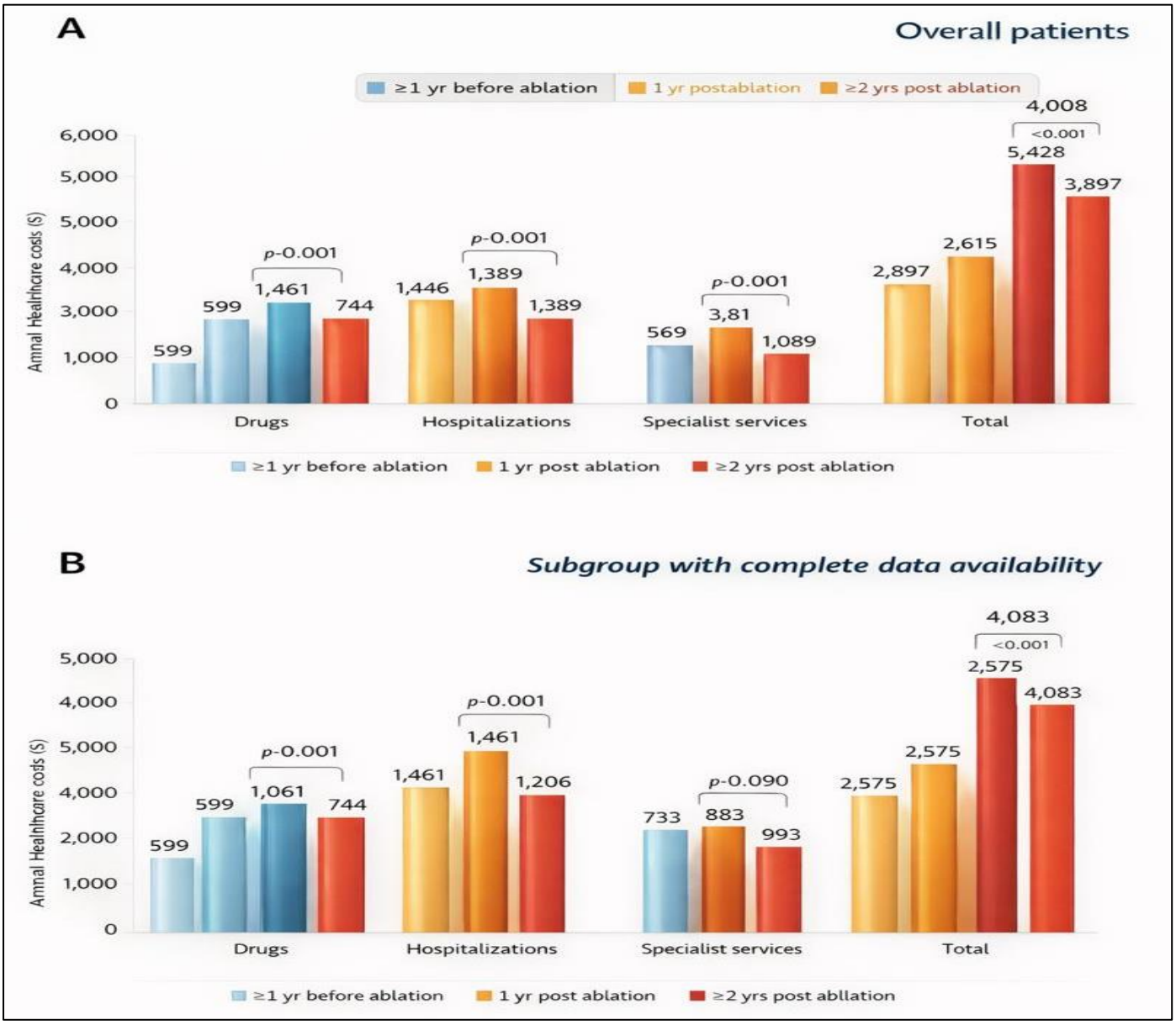

Fig 9 Temporal Changes in Healthcare Costs Before and After Ablation Across Patient Cohorts

These results demonstrate that clinical text is a critical driver of early detection performance, particularly at longer lead times, and that cross-modal attention provides a robust mechanism for integrating heterogeneous and partially observed data streams.

C. Model Interpretability and Clinical Plausibility

Beyond predictive accuracy, interpretability is essential for clinical adoption of early sepsis prediction systems. This section analyzes how the proposed multimodal model arrives at its predictions by examining attention mechanisms over time-series data and the relevance of textual concepts extracted from clinical notes. The goal is to demonstrate that model behavior aligns with established clinical reasoning rather than exploiting spurious correlations.

➢ *Attention Weight Analysis and Salient Temporal Patterns*

Temporal attention weights from the time-series encoder were analyzed to identify which physiological intervals most strongly influenced predictions. For septic patients, attention consistently concentrated on periods characterized by gradual but sustained deviations in vital signs and laboratory trends, rather than isolated extreme values.

Figure 10 illustrates an attention-based temporal learning pipeline that transforms longitudinal physiological signals into an interpretable early warning score. Multivariate vital signs and clinical indicators are first tracked over hours from admission and then segmented into fixed-length time steps for focused analysis. An attention heatmap highlights the relative importance of each physiological variable across time, revealing salient patterns associated with patient deterioration. These weighted temporal representations are ultimately aggregated to produce a high-confidence DEWS prediction, enabling proactive clinical intervention.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

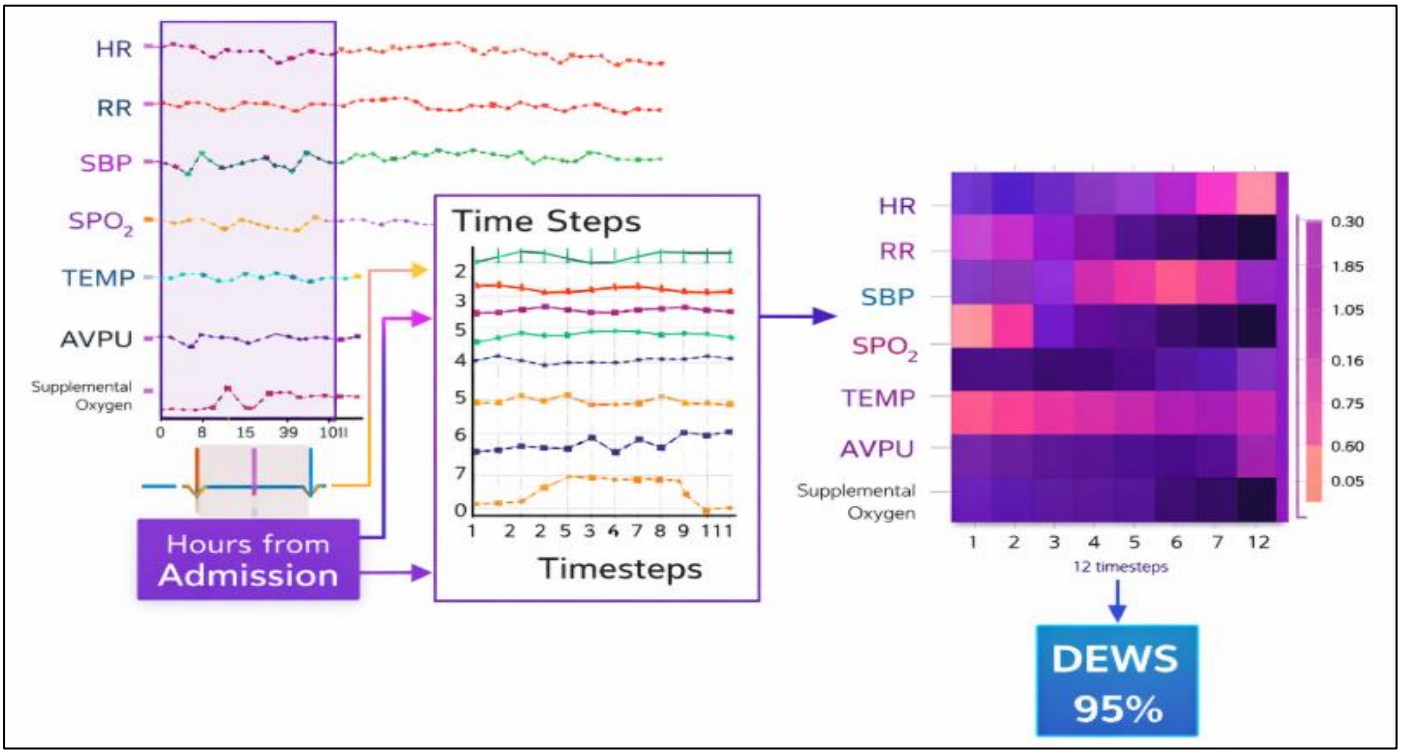

Fig 10 Attention-Driven Temporal Modeling of Physiological Trajectories for Deterioration Early Warning Scoring (DEWS)

Quantitatively, Table 6 summarizes the relative contribution of different variable groups, computed by aggregating normalized attention weights across patients.

Table 6 Average Attention Contribution by Variable Group

Variable Group	Mean Attention Weight (%)
Vital signs (HR, BP, RR, SpO_2)	38.4
Laboratory trends	31.7
Medication trajectories	18.9
Missingness & timing indicators	11.0

Vital signs and laboratory trends dominate early attention, reflecting their central role in detecting systemic instability. Medication-related features receive increasing attention closer to onset, consistent with escalation of therapy as clinicians respond to deterioration. The nontrivial weight assigned to missingness indicators supports the hypothesis that irregular measurement patterns themselves convey clinically relevant information.

Temporal analysis further reveals that attention shifts gradually forward in time as onset approaches, indicating that the model integrates cumulative evidence rather than reacting abruptly. This behavior aligns with clinical practice, where sepsis is often suspected after observing persistent trends rather than single abnormal readings.

➢ *Textual Concept Relevance and Clinical Alignment*

Interpretability of the text encoder was assessed by examining token- and segment-level attention scores. High-relevance text segments frequently contained concepts directly associated with early sepsis suspicion, including documentation of suspected infection, abnormal cultures, fever, hypotension, altered mental status, and escalation of antibiotics.

Table 7 lists the most frequently attended textual concept categories across septic cases, grouped by clinical theme.

Table 7 High-Impact Textual Concept Categories

Concept Category	Proportion of Septic Cases (%)
Suspected or confirmed infection	72.5
Antibiotic initiation or escalation	65.1
Hemodynamic instability	58.3
Fever or hypothermia	54.7
Altered mental status	41.9

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

These concepts are well aligned with established clinical reasoning for sepsis assessment, suggesting that the model leverages medically meaningful cues rather than superficial linguistic patterns. Importantly, many of these textual indicators appear in notes hours before formal sepsis diagnosis, explaining the model's improved performance at longer prediction horizons.

➢ *Clinical Plausibility and Trustworthiness*

Taken together, the attention analyses across modalities demonstrate that the model's predictions are driven by coherent physiological trends and clinically interpretable narrative cues. The alignment between salient features and established sepsis indicators supports the plausibility of the learned representations and provides a transparent basis for clinician review. By exposing when and why risk scores increase, the model offers interpretable early warnings that can be scrutinized and contextualized within routine clinical decision-making.

V. RECOMMENDATION AND CONCLUSION

➢ *Implications for Clinical Decision Support Systems*

The findings of this study have direct implications for the design and deployment of clinical decision support systems aimed at early sepsis detection. A multimodal prediction framework that integrates longitudinal physiological data with unstructured clinical narratives is well suited for real-time monitoring environments, where patient state evolves continuously and documentation occurs asynchronously. Integration into existing monitoring workflows can be achieved by embedding the model within electronic health record systems to operate on streaming vital signs, periodically updated laboratory results, and newly authored clinical notes. By generating updated risk scores at regular intervals, the system can provide clinicians with a dynamic view of sepsis risk rather than isolated alerts triggered at fixed thresholds.

Effective integration requires careful attention to workflow alignment. Alerts should be surfaced within interfaces already used by clinicians, such as patient dashboards or rounding tools, rather than through disruptive notification channels. Risk trajectories that show gradual escalation over time can support anticipatory decision-making, allowing clinicians to investigate potential infection sources, order confirmatory tests, or initiate closer monitoring before overt deterioration occurs. Presenting interpretability cues alongside risk scores, such as highlighted physiological trends or relevant note excerpts, can further support clinical understanding and situational awareness.

Balancing sensitivity and false alarm rates remains critical for clinician trust. While high sensitivity is essential to avoid missed cases of sepsis, excessive false positives can lead to alert fatigue and disengagement. The early-warning focus of the proposed approach enables a tiered alerting strategy, where low-confidence early signals prompt passive review and high-confidence signals trigger active alerts. Thresholds can be adjusted based on care setting, patient acuity, and staffing levels to ensure operational feasibility. Importantly, stable and temporally consistent predictions are more likely to be trusted than volatile alerts that fluctuate in response to minor data changes.

Ultimately, clinician trust depends not only on accuracy but also on transparency, reliability, and perceived clinical relevance. A decision support system that aligns with existing clinical reasoning, provides actionable lead time, and demonstrates consistent behavior across diverse care settings is more likely to be adopted and sustained. The multimodal approach presented in this work offers a foundation for such systems, supporting early intervention while respecting the cognitive and workflow constraints of frontline clinical practice.

➢ *Methodological Recommendations for Future Research*

Future research on early sepsis prediction should move beyond purely associative modeling toward approaches that explicitly address causality and uncertainty. While deep learning models are effective at identifying complex patterns, they do not distinguish between correlations and causal mechanisms underlying disease progression. Incorporating causal inference frameworks can help disentangle treatment effects, confounding by indication, and feedback loops introduced by clinical interventions. Methods such as causal graphs, counterfactual reasoning, and treatment-aware modeling can improve the reliability of predictions, particularly in settings where clinician actions influence both observed data and outcomes. Causal structure can also support more meaningful interpretability by clarifying whether rising risk estimates reflect disease evolution or responses to ongoing treatment.

Uncertainty estimation represents a complementary methodological priority. Early sepsis prediction inherently involves incomplete and noisy data, especially at long lead times. Future models should quantify predictive uncertainty alongside point estimates to support risk-aware decision-making. Techniques such as Bayesian deep learning, ensemble modeling, or calibrated confidence intervals can communicate when predictions are robust versus when they should be interpreted with caution. Explicit uncertainty signals may also enable adaptive alerting strategies, where clinicians are informed not only of elevated risk but also of the confidence associated with that assessment.

Extending multimodal models to account for multimorbidity is another critical direction. Many hospitalized patients present with multiple chronic conditions that alter baseline physiology, laboratory values, and documentation patterns. Models trained on single-disease paradigms may misinterpret deviations that are clinically appropriate for patients with complex comorbid profiles. Future research should explore representation learning strategies that encode long-term disease history and chronic condition interactions, allowing early sepsis predictors to adapt dynamically to heterogeneous patient phenotypes.

Finally, multi-task learning frameworks offer a promising avenue for improving generalization and clinical relevance. Rather than predicting sepsis in isolation, models

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

can be jointly trained to predict related outcomes such as acute organ dysfunction, need for vasopressors, ICU transfer, or in-hospital mortality. Shared representations learned across tasks can capture common physiological and narrative signals while reducing overfitting to any single endpoint. This approach aligns more closely with real-world clinical decision-making, where sepsis risk is assessed in the broader context of overall patient deterioration and resource utilization.

➢ *Conclusion*

This study presents a comprehensive multimodal deep learning framework for early sepsis prediction that integrates longitudinal clinical time series with unstructured medical text. By jointly modeling physiological signals, laboratory trends, medication trajectories, and narrative clinical documentation, the proposed approach addresses key limitations of traditional rule-based scores and unimodal predictive models. Quantitative results demonstrate consistent improvements in discrimination, precision under class imbalance, and clinically meaningful lead-time gains across multiple prediction horizons. Ablation and robustness analyses further show that cross-modal attention and hybrid fusion strategies play a central role in enabling early and stable detection, particularly when signals are sparse or asynchronous.

Beyond predictive performance, the study emphasizes interpretability and clinical plausibility. Attention analyses reveal that model predictions are driven by sustained physiological deterioration patterns and clinically relevant narrative cues, such as documentation of suspected infection or treatment escalation. These findings support the transparency and trustworthiness of the learned representations, reinforcing their potential suitability for clinical decision support. Cross-unit and cross-hospital evaluations indicate that the multimodal approach generalizes reasonably well across care settings while highlighting realistic challenges related to documentation variability, bias, and data shift.

In closing, multimodal deep learning represents a critical step toward proactive sepsis management. By providing earlier and more context-aware risk assessments, such systems can support timely intervention, improved resource allocation, and reduced morbidity and mortality. While further work is needed to incorporate causal reasoning, uncertainty estimation, and broader patient complexity, the framework outlined in this study establishes a strong methodological foundation for next-generation clinical decision support tools that move from reactive detection to anticipatory care.

REFERENCES

[1]. Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 72–78. https://doi.org/10.18653/v1/W19-1909

[2]. Ayoola, V. B., Idoko, I. P., Eromonsei, S. O., Afolabi, O., Apampa, A. R., & Oyebanji, O. S. (2024). The role of big data and AI in enhancing biodiversity conservation and resource management in the USA. *World Journal of Advanced Research and Reviews*, 23(2), 1851–1873.

[3]. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

[4]. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

[5]. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

[6]. Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5), 301–310. https://doi.org/10.1006/jbin.2001.1029

[7]. Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8, 6085. https://doi.org/10.1038/s41598-018-24271-9

[8]. Darko, D., Kwekutsu, E., & Idoko, I. P. (2025). Synergistic Effects of Phytochemicals in Combating Chronic Diseases with Insights into Molecular Mechanisms and Nutraceutical Development.

[9]. Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., … Das, R. (2016). Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Medical Informatics*, 4(3), e28. https://doi.org/10.2196/medinform.5909

[10]. Eguagie, M. O., Idoko, I. P., Ijiga, O. M., Enyejo, L. A., Okafor, F. C., & Onwusi, C. N. (2025). Geochemical and mineralogical characteristics of deep porphyry systems: Implications for exploration using ASTER. *International Journal of Scientific Research in Civil Engineering*, 9(1), 01–21.

[11]. Futoma, J., Hariharan, S., & Heller, K. (2017). Learning to detect sepsis with a multitask Gaussian process RNN classifier. *Proceedings of the 34th International Conference on Machine Learning*, 1174–1182.

[12]. Gaye, A., Victor Bamigwojo, O., Idoko, I. P., & Fatai Adeoye, A. (2025). Modeling Hepatitis B Virus Transmission Dynamics Using Atangana Fractional Order Network Approach: A Review of Mathematical and Epidemiological Perspectives. *International Journal of Innovative Science and Research Technology*, 10(4), 41–51.

[13]. Harutyunyan, H., Khachatrian, H., Kale, D. C., Steeg, G. V., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6, 96. https://doi.org/10.1038/s41597-019-0103-9

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan564

[14]. Henry, K. E., Hager, D. N., Pronovost, P. J., & Saria, S. (2015). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, 7(299), 299ra122. https://doi.org/10.1126/scitranslmed.aab3719

[15]. Idogho, C., Abah, E. O., Onuhc, J. O., Harsito, C., Omenkaf, K., Samuel, A., … Ali, U. E. (2025). Machine Learning-Based Solar Photovoltaic Power Forecasting for Nigerian Regions. *Energy Science & Engineering*, 13(4), 1922–1934.

[16]. Idoko, I. P., Akindele, J. S., Imarenakhue, W. U., & Bashiru, O. (2024). Exploring the Role of Bioenergy in Achieving Sustainable Waste Utilization and Promoting Low-Carbon Transition Strategies. *International Journal of Scientific Research in Science and Technology*. ISSN 2395-6011.

[17]. Idoko, I. P., Arthur, C., Ijiga, O. M., Osakwe, A., Enyejo, L. A., & Otakwu, A. (2024). Incorporating Radioactive Decay Batteries into the USA's Energy Grid: Solutions for Winter Power Challenges. *International Journal*, 3(9).

[18]. Idoko, I. P., David-Olusa, A., Badu, S. G., Okereke, E. K., Agaba, J. A., & Bashiru, O. (2024). The dual impact of AI and renewable energy in enhancing medicine for better diagnostics, drug discovery, and public health. *Magna Scientia Advanced Biology and Pharmacy*, 12(2), 99–127.

[19]. Idoko, I. P., Eniodunmo, O., Danso, M. O., Bashiru, O., Ijiga, O. M., & Manuel, H. N. N. (2024). Evaluating benchmark cheating and the superiority of MAMBA over transformers in Bayesian neural networks: An in-depth analysis of AI performance. *World Journal of Advanced Engineering Technology and Sciences*, 12(1), 372–389.

[20]. Idoko, I. P., Ijiga, O. M., Akoh, O., Agbo, D. O., Ugbane, S. I., & Umama, E. E. (2024). Empowering sustainable power generation: The vital role of power electronics in California's renewable energy transformation. *World Journal of Advanced Engineering Technology and Sciences*, 11(1), 274–293.

[21]. Idoko, I. P., Ijiga, O. M., Enyejo, L. A., Akoh, O., & Isenyo, G. (2024). Integrating superhumans and synthetic humans into the Internet of Things (IoT) and ubiquitous computing: Emerging AI applications and their relevance in the US context. *Global Journal of Engineering and Technology Advances*, 19(01), 006–036.

[22]. Idoko, I. P., Ijiga, O. M., Harry, K. D., Ezebuka, C. C., Ukatu, I. E., & Peace, A. E. (2024). Renewable energy policies: A comparative analysis of Nigeria and the USA. *World Journal of Advanced Research and Reviews*, 21(1), 888–913.

[23]. Idoko, I. P., Ayodele, T. R., Abolarin, S. M., & Ewim, D. R. E. (2023). Maximizing the cost effectiveness of electric power generation through the integration of distributed generators: wind, hydro and solar power. *Bulletin of the National Research Centre*, 47(1), 166.

[24]. Ijiga, A. C., Abutu, E. P., Idoko, P. I., Ezebuka, C. I., Harry, K. D., Ukatu, I. E., & Agbo, D. O. (2024). Technological innovations in mitigating winter health challenges in New York City, USA. *International Journal of Science and Research Archive*, 11(01), 535–551.

[25]. Ijiga, A. C., Peace, A. E., Idoko, I. P., Agbo, D. O., Harry, K. D., Ezebuka, C. I., & Ukatu, I. E. (2024). Ethical considerations in implementing generative AI for healthcare supply chain optimization: A cross-country analysis across India, the United Kingdom, and the United States of America. *International Journal of Biological and Pharmaceutical Sciences Archive*, 7(01), 048–063.

[26]. Ijiga, O. M., Idoko, I. P., Enyejo, L. A., Akoh, O., & Ileanaju, S. (2024). Harmonizing the voices of AI: Exploring generative music models, voice cloning, and voice transfer for creative expression. *World Journal of Advanced Engineering Technology and Sciences*, 11, 372–394.

[27]. Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention. *Journal of Science and Technology*, 11, 001–024.

[28]. Ikedionu, C. A., Idoko, I. P., Omale, J. O., & Idogho, C. (2025). Mathematical modeling of 3D printing of microreactors for continuous flow chemical processes. *International Journal of Research Publication and Reviews*.

[29]. Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., … Cheang, M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*, 34(6), 1589–1596. https://doi.org/10.1097/01.CCM.0000217961.75225.E9

[30]. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682

[31]. Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2016). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*.

[32]. Manuel, H. N. N., Adeoye, T. O., Idoko, I. P., Akpa, F. A., Ijiga, O. M., & Igbede, M. A. (2024). Optimizing passive solar design in Texas green buildings by integrating sustainable architectural features for maximum energy efficiency. *Magna Scientia Advanced Research and Reviews*, 11(01), 235–261.

[33]. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6, 26094. https://doi.org/10.1038/srep26094

[34]. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. *Proceedings of the 28th International Conference on Machine Learning*, 689–696.

[35]. Okika, N., Nwatuzie, G. A., Odozor, L., Oni, O., & Idoko, I. P. (2025). Addressing iot-driven

cybersecurity risks in critical infrastructure to safeguard public utilities and prevent large-scale service disruptions. *International Journal of Innovative Science and Research Technology*, 10(2).

[36]. Pakhomov, S. V. S., Finley, G., McEwan, R., Wang, Y., & Melton, G. B. (2010). Corpus domain effects on distributional semantic modeling of medical terms. *Bioinformatics*, 26(24), 2991–2998. https://doi.org/10.1093/bioinformatics/btq567

[37]. Raghu, A., Komorowski, M., Celi, L. A., Szolovits, P., & Ghassemi, M. (2017). Continuous state-space models for optimal sepsis treatment: A deep reinforcement learning approach. *Machine Learning for Healthcare Conference Proceedings*, 77, 147–163.

[38]. Raith, E. P., Udy, A. A., Bailey, M., McGloughlin, S., MacIsaac, C., Bellomo, R., & Pilcher, D. V. (2017). Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. *JAMA*, 317(3), 290–300. https://doi.org/10.1001/jama.2016.20328

[39]. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1, 18. https://doi.org/10.1038/s41746-018-0029-1

[40]. Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., … Naghavi, M. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *The Lancet*, 395(10219), 200–211. https://doi.org/10.1016/S0140-6736(19)32989-7

[41]. Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5), 507–513. https://doi.org/10.1136/jamia.2009.001560

[42]. Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., … Angus, D. C. (2016). Assessment of clinical criteria for sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 762–774. https://doi.org/10.1001/jama.2016.0288

[43]. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063

[44]. Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., … Angus, D. C. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801–810. https://doi.org/10.1001/jama.2016.0287

[45]. Ugbane, S. I., Umeaku, C., Idoko, I. P., Enyejo, L. A., Michael, C. I., & Efe, F. (2024). Optimization of quadcopter propeller aerodynamics using blade element and vortex theory. *International Journal of Innovative Science and Research Technology*, 9(10), 1–12.

[46]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.