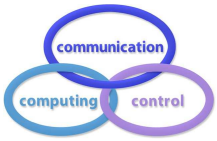
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 3, Month: June, Year: 2025
Article Number: 6968, https://doi.org/10.15837/ijccc.2025.3.6968



CCC Publications

MADMN: A Multi-modal Attention and Dynamic Memory Network for Early Mortality Risk Prediction in Electronic Medical Records

Yufei Chen, Kai Chen

Yufei Chen
Changzhou Saixun Network Technology Co., Ltd., Jiangsu, China
cyf2000404@163.com

Kai Chen*
1. Changzhou No.2 People's Hospital, Jiangsu, China
2. The Third Affiliated Hospital of Nanjing Medical University, Changzhou 213003, Jiangsu, China
*Corresponding author: c4kaichen@163.com

Abstract

This paper proposes a novel Multi-modal Attention and Dynamic Memory Network (MADMN) model for early mortality risk prediction based on Electronic Medical Records (EMRs). The model integrates multi-modal feature extraction, cross-modal attention fusion, and dynamic memory networks within a unified framework to process structured, time-series, and textual data. MADMN effectively captures complex temporal dependencies and multimodal interactions, enhancing prediction accuracy and interpretability. Experimental results demonstrate that MADMN significantly outperforms traditional machine learning and deep learning baselines in terms of Accuracy, F1 Score, and ROC-AUC. Furthermore, SHAP analysis validates the model's interpretability by highlighting key features contributing to predictions. The model also supports counterfactual analysis, enabling personalized treatment decisions and resource optimization. MADMN offers a robust and interpretable solution for multimodal medical data analysis and risk prediction, paving the way for advancements in precision medicine.

Keywords: Multimodal Data Analysis, Dynamic Memory Network, Cross-modal Attention, Electronic Medical Records, Risk Prediction

1 Introduction

Electronic Medical Records (EMRs) have become a core component of modern healthcare systems, providing critical support for patient information management and clinical decision-making. EMRs store a wide range of data, including demographic information, laboratory test results, diagnostic reports, and treatment histories. They also contain multimodal data such as medical images and clinical notes, which are instrumental in disease diagnosis, treatment planning, and risk prediction. Particularly in intensive care settings, EMR-based risk prediction systems play a vital role in identifying high-risk patients and enabling early interventions, thereby reducing mortality rates and optimizing healthcare resource allocation.

The rapid advancement of machine learning and deep learning technologies has significantly improved the accuracy and efficiency of EMR data analysis, enabling the processing of complex medical data that was previously infeasible [1][2][3]. However, the multimodal and dynamic characteristics of EMR data present considerable challenges for modeling and interpretability. One of the primary challenges in risk prediction lies in the complexity of multimodal data. EMRs typically consist of structured data, time-series data, and unstructured text, each with distinct statistical properties and analytical requirements. For instance, laboratory test results are often stored in structured formats, physiological monitoring data exhibit temporal dynamics, and clinical notes are recorded as unstructured free text. Effectively integrating these heterogeneous data types to capture their latent relationships and complementary information is critical for improving model performance. Additionally, diseases often progress dynamically, and patients' health statuses evolve over time. This necessitates models that not only capture long-term dependencies but also reflect short-term fluctuations to accurately predict disease progression.

In addition to data complexity, a lack of model interpretability poses another major challenge. Although deep learning models have demonstrated remarkable performance in risk prediction tasks, their complex and opaque structures—often referred to as "black boxes"—limit clinicians' ability to understand and trust the predictions [4] [5] [6]. In clinical applications, decision-making depends not only on predictive accuracy but also on the ability to explain the rationale behind predictions, supporting transparent and informed diagnoses and treatment plans. Consequently, developing risk prediction models with interpretable outputs has emerged as a critical research priority. Moreover, current methods often lack sufficient capabilities for modeling interactions among multimodal data and extracting dynamic features, leading to limitations in prediction comprehensiveness and adaptability.

Despite the progress made in EMR-based risk prediction, several limitations remain. First, many existing models focus on analyzing single data modalities, such as structured data or time-series data, neglecting the interactions between modalities. For example, neural network models based solely on structured data may perform well for specific tasks but fail to leverage potential insights from text and time-series data [7] [8] . Second, while some approaches incorporate dynamic modeling, their ability to capture temporal dependencies remains limited. Traditional regression models and static machine learning methods are inherently incapable of modeling temporal dynamics, and although recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have improved upon this, they still lack mechanisms for effectively integrating multimodal data, thereby constraining their predictive performance[9].

Moreover, poor interpretability limits the clinical applicability of current models. Although some studies have explored anomaly detection and medical data processing techniques using deep learning, these methods often lack transparency, leaving clinicians unable to comprehend the reasoning behind model decisions. This lack of interpretability is particularly problematic in high-risk scenarios, such as cardiovascular disease prediction, where high accuracy alone is insufficient to meet clinical demands for reliability and explainability.

To address these challenges, this paper proposes a novel model that integrates a cross-modal attention mechanism with a dynamic memory network (DMN) to overcome the shortcomings of existing approaches. First, the model dynamically learns interactions between multimodal data through cross-modal attention, effectively leveraging complementary information from structured data, time-series data, and text data. Second, it models the temporal evolution of patients' conditions using a dynamic memory network, capturing both long-term dependencies and short-term fluctuations to enhance prediction performance and time sensitivity. Additionally, interpretable analysis tools are incorporated to improve model transparency and clinical usability.

The proposed approach offers several key advantages. First, it supports multimodal feature fusion within a unified analytical framework, enabling the integration of structured, time-series, and textual data to exploit their complementary relationships. Second, the model dynamically identifies critical features and adapts to complex disease progression patterns, improving its ability to process temporal dynamics. Furthermore, by providing interpretable analysis alongside risk predictions, the model enhances transparency and reliability, helping clinicians understand the predictions and make well-informed decisions. These features make the model particularly suitable for real-world medical

applications.

The main contributions of this paper are as follows: (1) It designs a novel model that combines a cross-modal attention mechanism and a dynamic memory network to analyze multimodal data and predict dynamic risks; (2) It provides a unified framework for modeling multimodal data, addressing challenges related to data fusion and temporal dynamics; and (3) It demonstrates, through extensive experiments, that the proposed model achieves superior performance in mortality risk prediction tasks, offering high accuracy and strong interpretability for clinical decision support.

2 Related Works

EMRs contain diverse types of data that collectively reflect a patient's health status and disease progression. To build comprehensive risk prediction models, it is essential to effectively utilize the complementary information embedded in these data modalities. EMR data can generally be categorized into three primary types: image data, time-series and trajectory data, and structured and text data. Each modality serves a distinct purpose in clinical analysis. Image data, such as CT and MRI scans, provide visual representations of anatomical structures and pathological changes, enabling direct identification of lesions or abnormalities. Time-series and trajectory data capture dynamic changes in physiological signals over time, offering insights into disease progression and treatment responses. Structured and text data, including demographic information, diagnostic codes, laboratory results, and clinical notes, contain detailed patient histories and contextual information that support decision-making. By reviewing these three data modalities, this section aims to highlight their respective contributions to EMR-based risk prediction and discuss existing approaches for their integration and analysis. This categorization also lays the foundation for understanding the challenges and opportunities in multi-modal data fusion, as addressed in subsequent sections.

2.1 Image-Based Risk Prediction

Image data analysis has become a critical approach in electronic medical record (EMR)-based risk prediction. Leveraging the outstanding performance of deep learning and machine learning models, significant progress has been made in medical image analysis, particularly in disease classification, lesion detection and segmentation, and model optimization. Recent studies have focused on predicting prognosis and molecular classifications of cancers. For example, Foersch et al. [10] and Fremond et al. [11] employed deep learning techniques to analyze cancer images, effectively predicting colorectal and endometrial cancer outcomes. These studies highlight the potential of deep learning for complex image analysis, offering novel methods for disease risk prediction and personalized treatment planning.

In lesion detection and segmentation, Saeedi et al. [12] proposed a convolutional neural network (CNN)-based model that successfully identified the location and size of brain tumors, providing insights for early detection and intervention. Similarly, Curila et al. [13] combined statistical methods with edge-based segmentation techniques to process CT images of COVID-19 infections, enhancing the precision of lesion detection. These studies demonstrate the potential of integrating deep learning with traditional image processing techniques, particularly for complex lesion detection tasks. Furthermore, Mahmood et al. [14] utilized MRI images to detect acute knee injuries, offering new technical support for sports injury diagnosis and prediction.

Recent research has also focused on model optimization and performance improvement. Vrbančič et al. [15] optimized CNN hyperparameters, significantly improving the accuracy of COVID-19 classification from X-ray images. Singh et al. [16] explored explainable deep learning approaches to enhance transparency and interpretability in medical image retrieval, providing higher trustworthiness for clinical applications. Additionally, Babu et al. [17] proposed an optimized CNN-based approach for brain tumor segmentation and classification, leveraging artificial bee colony optimization and thresholding techniques to improve segmentation accuracy and computational efficiency. Similarly, Lee et al. [18] applied AI-based abdominal CT measurements to predict mortality and cardiometabolic disease risk, demonstrating the potential of AI in improving disease risk assessment through advanced imaging analysis. These studies highlight ongoing efforts to refine machine learning models, optimizing them for better clinical applicability and prediction accuracy.

Despite these advances, several challenges remain. First, the acquisition and annotation of medical imaging data are costly, resulting in limited training datasets and reduced model generalizability. Second, although explainable artificial intelligence techniques have been developed, most deep learning models still lack sufficient transparency, limiting the trust clinicians place in their predictions. Moreover, multimodal data integration is relatively underexplored, as most studies rely solely on single-modal image data without incorporating text, time-series, or genomic data available in EMRs. This limits the comprehensive performance of prediction models. Finally, the lack of large-scale clinical validation affects the generalizability and applicability of these methods. While many models perform well in experimental settings, their adaptability and stability in real-world environments require further evaluation.

2.2 Sequence-Based Risk Prediction

Time-series and trajectory data analysis play a vital role in EMR-based risk prediction by capturing patients' evolving health conditions over time. These methods uncover disease progression patterns and provide scientific support for early detection and personalized interventions. Advances in deep learning and machine learning technologies have further improved the accuracy and applicability of time-series data analysis.

Placido et al. [19] developed a deep learning algorithm to analyze patient trajectories and predict pancreatic cancer risk. By leveraging long-term EMR data, the model identified critical milestones in disease progression, enabling early detection and intervention. This study emphasizes the importance of trajectory data for understanding complex diseases and demonstrates the ability of machine learning models to process high-dimensional time-series data effectively.

Forrest et al. [20] further validated the potential of time-series data. They developed a prediction model for coronary artery disease using longitudinal data and tested it on two long-term follow-up cohorts. The study showcased the practical value of time-series analysis for chronic disease monitoring and highlighted the importance of long-term data accumulation for training risk prediction models. Similarly, Morid et al. [9] explored time-series forecasting methods, employing recurrent neural networks (RNNs) and long short-term memory (LSTM) networks to capture complex temporal dependencies, achieving improved stability and reliability in predictions.

Beyond disease risk prediction, time-series analysis has demonstrated effectiveness in chronic disease management and real-time monitoring. For example, Haque et al. [4] employed enhanced neural networks to analyze diagnostic records and successfully predict asthma onset risks. Yuantai [6] proposed an intelligent wearable system based on a discrete chaotic fuzzy neural network to monitor health status and activity patterns in real time, providing a new technological direction for personalized health management. Beyond disease risk prediction, time-series analysis has demonstrated effectiveness in chronic disease management and real-time monitoring. For example, Haque et al. [4] employed enhanced neural networks to analyze diagnostic records and successfully predict asthma onset risks. Yuantai [6] proposed an intelligent wearable system based on a discrete chaotic fuzzy neural network (DC-FNN) to monitor health status and activity patterns in real time, providing a new technological direction for personalized health management. Renc et al. [21] introduced a zero-shot health trajectory prediction model using transformers, a novel approach that leverages the power of deep learning to predict future health trajectories without the need for labeled training data, offering new possibilities for personalized healthcare. Yang et al. [22] applied deep reinforcement learning to optimize fairness and bias control in machine learning models, improving clinical decision support stability and adaptability. Additionally, Aminizadeh et al. [7] investigated distributed computing and IoT-based frameworks, demonstrating the potential of time-series data in large-scale medical networks.

Despite these achievements, challenges remain. EMR time-series data often suffer from missing values, inconsistent records, and noise, which pose threats to model training and prediction accuracy. Although deep learning models excel at capturing complex temporal dependencies, their lack of interpretability continues to limit clinical applications. Transparent algorithms are needed to help clinicians and patients understand the logic behind predictions. Furthermore, multimodal data integration remains underdeveloped, as most studies focus on single-type time-series data while overlooking text, imaging, and genomic data, constraining the models' comprehensiveness in risk evaluation.

2.3 Text-Based Risk Prediction

Structured and text data analysis represents an essential research direction in EMR-based risk prediction, focusing on extracting and analyzing demographic information, diagnostic codes, laboratory results, and clinical notes. These data contain detailed records of patients' medical histories and treatment processes. By leveraging machine learning and natural language processing (NLP) techniques, valuable information can be mined to support disease risk prediction and personalized health management.

Feng et al. [8] developed an automated ICD-11 coding model that transforms medical text into structured classification data, improving EMR management efficiency. This approach reduces manual coding workload while enhancing prediction reliability through standardized processing. Similarly, Kim et al. [5] proposed a deep learning-based algorithm to detect anomalies and manipulations in medical data, strengthening EMR security and integrity and ensuring the reliability of data analysis.

For risk prediction, Pfob et al. [23] built a machine learning model that combined patient-reported outcomes with structured medical data to predict one-year follow-up outcomes after breast cancer surgery. This approach integrates subjective feedback with objective medical data, offering new insights into post-operative risk assessment. Oh [3] explored big data analytics for hyperlipidemia diagnosis, leveraging structured diagnostic test results and laboratory data to improve early disease identification. Yashudas et al. [24] proposed a cardiovascular disease prediction system based on IoT networks, combining structured EMR data with real-time monitoring to provide dynamic risk assessments and personalized treatment recommendations.

In distributed data processing and multi-source data integration, Aminizadeh et al. [7] investigated distributed computing and IoT frameworks for medical data processing, highlighting their potential for remote monitoring and intelligent health management. Haque et al. [4] demonstrated the practicality of structured data in chronic disease management by predicting asthma risks using advanced neural network algorithms. Moreover, Wan and Tian [25] proposed a novel machine learning approach for user stress detection using social media text, opening new possibilities for integrating external data sources into health risk prediction systems. Similarly, Zhu et al. [26] identified proteomic signatures of healthy dietary patterns associated with lower risks of chronic diseases and mortality, further underscoring the importance of integrating diverse data types in comprehensive health risk prediction.

Despite the potential of structured and text data analysis, several challenges remain. Text data often lack standardization, and semantic understanding depends heavily on context and medical terminology, posing high demands on NLP models. Moreover, EMR data frequently contain missing values and redundant information, affecting model training and predictive performance. Additionally, large-scale text data processing requires substantial computational resources, and increasing model complexity often compromises interpretability, potentially limiting clinical adoption.

2.4 Literature Review Summary

Recent studies have made significant progress in leveraging EMR data for disease risk prediction, including image analysis, time-series modeling, and structured and text data interpretation. Image data analysis has shown promise in disease classification and lesion detection, but challenges such as high data acquisition costs, limited interpretability, and insufficient integration with other data types remain barriers to broader applicability [27]. Time-series modeling effectively captures temporal dependencies and disease trajectories, yet issues like data quality, missing values, and poor interpretability restrict clinical adoption [28, 29]. Structured and text data provide rich contextual information for risk prediction but face challenges related to standardization, semantic ambiguity, and computational complexity in large-scale processing [30].

Despite these advancements, limitations persist. Many methods focus on single-modality data, overlooking interactions between modalities. Multimodal data fusion, which integrates visual, temporal, and textual features, is still underdeveloped. Additionally, while deep learning models achieve high predictive accuracy, their lack of interpretability reduces clinical usability [31].

3 Methodology

3.1 Overall Framework

This paper proposes an innovative Multi-modal Attention and Dynamic Memory Network (MADMN) for early mortality risk prediction based on EMRs. MADMN integrates multi-modal feature extraction, cross-modal attention mechanisms, and DMNs into a unified modeling framework. It effectively processes heterogeneous medical data while capturing the dynamic evolution of a patient's condition over time. Moreover, the model enhances its representation capacity for complex clinical data through dynamic interactions among multi-modal features, thereby improving prediction accuracy and interpretability.

The overall framework consists of four core modules: multi-modal feature extraction, cross-modal attention fusion, dynamic memory modeling, and risk prediction output. Multi-modal data is first processed using specialized techniques for each modality—structured features are extracted using a variational autoencoder (VAE), time-series data is modeled through a time-aware transformer, and textual information is embedded using ClinicalBERT. These features are then integrated via a cross-modal attention mechanism that dynamically highlights important relationships across modalities. The dynamic memory network captures temporal dependencies and short-term fluctuations, providing a refined representation for risk prediction. Finally, a fully connected layer outputs mortality risk scores, enabling accurate and interpretable assessments.

Key innovations of MADMN include its ability to integrate multi-modal data through cross-modal attention and dynamically model temporal patterns using memory networks, distinguishing it from prior works. It also enhances interpretability through SHAP analysis and counterfactual reasoning, offering actionable insights for clinical decision-making. Figure 1 provides a visual overview of the framework to illustrate the data flow and interactions between modules. This design improves clarity and highlights MADMN's unique contributions compared to existing models.

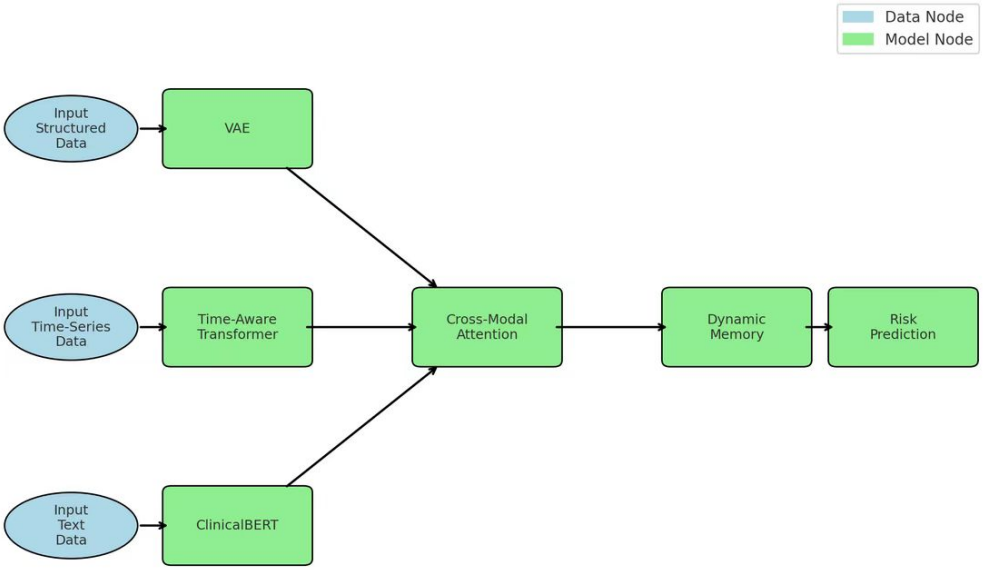


Figure 1: Neural Network Structure for MADMN Framework.

3.2 Multi-Modal Feature Extraction

The multi-modal feature extraction module is designed to process three primary types of data in EMRs: structured data, time-series data, and text data. Structured data, such as demographic information, laboratory test results, and diagnostic codes, captures quantitative aspects of a patient's condition. However, it often suffers from high dimensionality, noise, and missing values. To address these issues, a Variational Autoencoder (VAE) is employed to extract latent features while simultaneously denoising the input. Given structured data input X_{struct}, the VAE maps it to a latent variable

representation z that follows a Gaussian distribution:

$$z \sim q\left(z \mid X_{\text{struct}}\right) \tag{1}$$

The decoder then reconstructs the input based on the latent variable, enabling the model to learn meaningful representations while reducing noise. The encoded structured feature representation is denoted as:

$$H_{\text{struct}} = VAE\left(X_{\text{struct}}\right) \tag{2}$$

To further enhance the robustness and stability of the model, a residual connection is incorporated. This approach ensures gradient flow during backpropagation, improving convergence and feature representation:

$$H_{\text{struct}}^{'} = H_{\text{struct}} + ReLU\left(W_{s1}H_{\text{struct}} + b_{s1}\right) \tag{3}$$

Here, W_{s1} and b_{s1} are learnable parameters, and ReLU serves as a nonlinear activation function. This design prevents vanishing gradients and strengthens the expressive power of structured data features, ensuring more informative input representations for downstream tasks.

Time-series data, such as heart rate, blood pressure, and oxygen saturation, provides critical insights into physiological changes and disease progression. It exhibits dynamic patterns over time, necessitating temporal modeling techniques. A Time-Aware Transformer is employed to capture both short-term fluctuations and long-term dependencies. The modeling begins with a self-attention mechanism, which computes relationships between different time steps through the query (Q), key (K), and value (V) matrices:

$$H_{\text{seq}} = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{4}$$

where d represents the feature dimension. The self-attention mechanism dynamically adjusts the importance of each time step, enabling the model to highlight influential moments in the temporal sequence.

Since transformers lack inherent sequence order information, positional encoding is added to introduce temporal order. The encoding is defined as:

$$\begin{aligned} PE_{(pos,2i)} &= \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right), \\ PE_{(pos,2i+1)} &= \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right), \end{aligned} \tag{5}$$

where pos denotes the position index in the sequence, and d_{model} is the embedding dimension. By combining sine and cosine functions with varying frequencies, the positional encoding ensures that each time step has a unique representation while preserving relative ordering information. This enables the model to capture periodic trends and long-term dependencies, making it suitable for irregularly sampled and temporally evolving time-series data.

Text data, including clinical notes, medical histories, and examination reports, provide rich contextual and semantic information. To process this unstructured data, the model employs ClinicalBERT, a domain-specific variant of BERT pre-trained on biomedical corpora. ClinicalBERT can effectively capture contextual dependencies and medical terminologies, making it well-suited for modeling complex medical text. Given an input text sequence X_{text}, the embeddings are computed as follows:

$$H_{\text{text}} = BERT\left(X_{\text{text}}\right) \tag{6}$$

where H_{text} represents contextual embeddings generated by ClinicalBERT. These embeddings encode not only word-level meanings but also the relationships among words based on surrounding context.

The embedding process in ClinicalBERT includes three components: Token Embedding: Assigns vector representations to individual words or subwords. Segment Embedding: Distinguishes between different segments in a document or sentence pairs. Position Embedding: Captures the order of words in a sequence to retain sentence structure.

By passing these embeddings through multiple transformer layers, ClinicalBERT produces context-aware representations that capture semantic and syntactic patterns in clinical text. For instance, terms like "hypertension" and "high blood pressure" are recognized as closely related, improving performance in downstream tasks such as mortality risk prediction.

The integration of structured, time-series, and text data creates a unified representation that leverages complementary information from different modalities. Structured data provides quantitative patterns, time-series data models temporal evolution, and text data offers contextual knowledge, enabling the model to capture complex relationships in clinical scenarios.

To summarize, the multi-modal feature extraction module processes structured, temporal, and textual data using dedicated architectures—VAE for structured data, Time-Aware Transformer for time-series data, and ClinicalBERT for text data. This design ensures that each data type is effectively represented and provides a solid foundation for downstream tasks such as prediction and decision support. The combination of these techniques not only reduces the impact of data heterogeneity but also improves robustness and interpretability in clinical applications.

3.3 Cross-Modal Attention Fusion

The cross-modal attention fusion module is a critical component of the model, designed to address the heterogeneity and interdependence of multi-modal data in EMRs. Directly concatenating features from different modalities often results in information loss and fails to capture the intricate interactions between modalities. To overcome this limitation, the cross-modal attention mechanism dynamically learns the relationships between features from various modalities and integrates them into a unified representation. This mechanism allows the model to highlight and prioritize important features while downplaying less relevant ones, thus improving the overall predictive performance.

The fusion process begins by computing the attention weights between features from different modalities. For a query matrix Q_i derived from modality i and a key matrix K_j derived from modality j, the attention weight A_{ij} is computed as:

$$A_{ij} = Softmax\left(\frac{Q_i K_j^T}{\sqrt{d}}\right) V_j \tag{7}$$

Here, Q_i, K_j, V_j represent the query, key, and value matrices of modalities i and j, respectively, and d is the dimensionality of the feature space. The softmax function ensures that the attention weights are normalized and sum to 1, enabling the model to selectively focus on relevant features across modalities.

Once the attention weights are computed, the features from all modalities are aggregated into a fused representation using a dynamic weighting scheme. The fused representation H_{fused} is expressed as:

$$H_{\text{fused}} = \alpha_1 H_{\text{struct}} + \alpha_2 H_{\text{seq}} + \alpha_3 H_{\text{text}} \tag{8}$$

where $H_{\text{struct}}, H_{\text{seq}}$ and H_{text} denote the features extracted from structured, time-series, and text data, respectively. The weights $\alpha_1, \alpha_2, \alpha_3$ are dynamically computed based on the importance of each modality. These weights are generated through a learnable softmax layer applied to the concatenated features:

$$\alpha = Softmax\left(W H_{\text{cross}}\right) \tag{9}$$

Here, W is a learnable weight matrix, and H_{cross} is the concatenation of all modality-specific features. The softmax operation ensures that the weights are non-negative and sum to 1, allowing the model to adaptively adjust the contribution of each modality depending on the task and data characteristics.

To further enhance the robustness of the fusion process, a sparsity constraint is applied to the attention matrix A_{ij}. The sparsity regularization term encourages the model to focus on a limited number of key features, thus avoiding overfitting and improving interpretability:

$$L_{\text{attu}} = \| A \|_1 \tag{10}$$

By penalizing the sum of the absolute values of the attention weights, this constraint ensures that only the most relevant features are given significant attention during the fusion process. The cross-modal attention fusion mechanism not only integrates multi-modal features but also facilitates information sharing between modalities. For instance, trends in time-series data (e.g., a sudden drop in blood pressure) can be linked to specific diagnostic codes in structured data or textual descriptions (e.g., "patient experiencing hypotension"). This ability to model interactions and dependencies across modalities allows the model to capture richer and more informative representations, ultimately leading to improved performance in downstream tasks.

In summary, the cross-modal attention fusion module transforms the heterogeneous information from structured, time-series, and text data into a unified and contextually enriched representation. By dynamically adjusting the importance of each modality and focusing on key features, this module lays the foundation for effective downstream modeling and decision-making in clinical settings.

3.4 DMN Modeling

The DMN is designed to model temporal dynamics in multi-modal clinical data, capturing both short-term changes and long-term trends in a patient's condition. Unlike static models, the DMN maintains a memory state that evolves over time, enabling it to reflect the progression of a patient's physiological status or disease condition. The DMN begins by initializing a memory state based on the fused multi-modal feature representation H_{fused}, obtained from the cross-modal attention fusion module. The initialization is defined as:

$$M_0 = f\left(H_{\text{fused}}\right) \tag{11}$$

where $f(\cdot)$ is a learnable transformation, such as a fully connected layer with a non-linear activation function. This initial memory state encodes the patient's baseline condition and serves as the starting point for temporal modeling.

At each time step t, the memory state is updated using a gated recurrent unit (GRU). The GRU integrates new information from the current fused feature representation $H_{\text{fused,t}}$ while retaining relevant historical context from the previous memory state M_{t-1}. The memory update process is defined as:

$$M_t = GRU\left(M_{t-1}, H_{\text{fused,t}}\right) \tag{12}$$

This update mechanism allows the DMN to dynamically adjust its focus, emphasizing features that are most relevant at each time step. The GRU's gating structure ensures that only significant new information is incorporated, preventing noise or redundant data from overwhelming the model.

To refine the patient representation further, the DMN uses an attention mechanism to highlight key features at each time step. The attention mechanism computes a weighted combination of the fused features H_{fused}, guided by the current memory state M_t, and generates the final output representation:

$$H_{\text{out}} = Attention\left(M_t, H_{\text{fused}}\right) \tag{13}$$

This iterative reasoning process ensures that the model can focus on clinically significant patterns, such as rapid changes in vital signs, emerging symptoms, or notable events described in clinical notes. By dynamically interacting with multi-modal features, the DMN captures complex dependencies and contextual information critical for understanding a patient's evolving condition.

The DMN's ability to maintain and update a memory state, combined with its attention-based reasoning, makes it well-suited for clinical prediction tasks. It provides a comprehensive temporal view of a patient's status while ensuring that the most relevant features are emphasized, improving both the accuracy and interpretability of predictions in real-world clinical applications.

3.5 Risk Prediction Module

The risk prediction module is the final stage of the model, responsible for generating the predicted mortality risk score based on the multi-modal and temporal representations derived from the DMN. This module consolidates the learned features into a prediction that reflects the likelihood of an adverse outcome, such as mortality, for a given patient.

The output of the DMN, H_{out}, serves as the input to the risk prediction module. A fully connected layer is used to map this high-dimensional representation to a single scalar value or a set of probabilities, depending on the nature of the prediction task. In the case of binary classification, the model outputs the predicted probability of the positive class (e.g., mortality):

$$\widehat{y} = \sigma\left(WH_{\text{out}} + b\right) \tag{14}$$

where W and b are learnable parameters of the fully connected layer, and $\sigma(\cdot)$ is the sigmoid activation function that maps the output to a probability in the range $[0, 1]$.

To train the model, a binary cross-entropy loss function is employed, which measures the difference between the predicted probability $\widehat{y}$ and the ground truth label y. The loss function is defined as:

$$L_{\text{risk}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\widehat{y}_i\right) + \left(1 - y_i\right)\log\left(1 - \widehat{y}_i\right)\right] \tag{15}$$

Here, $y_i \in \{0, 1\}$ represents the ground truth label for the i-th patient, $\widehat{y}_i$ is the predicted probability for the same patient and N is the total number of patients in the training dataset.

The binary cross-entropy loss function penalizes incorrect predictions and rewards accurate ones, pushing the model to minimize the overall error. By backpropagating this loss, the model learns to adjust its parameters to optimize prediction performance.

The risk prediction module ensures that the model outputs interpretable and actionable results. For example, the predicted probability can be used to stratify patients into risk categories (e.g., low, medium, high) or to trigger early interventions for high-risk individuals. By combining multi-modal feature extraction, cross-modal attention, and temporal modeling, the module provides a robust foundation for clinical decision-making in real-world scenarios.

4 Experimental Results

4.1 Dataset

This study utilizes the MIMIC-III (Multiparameter Intelligent Monitoring in Intensive Care) dataset, which contains over 60,000 ICU admission records. The dataset includes a rich collection of medical data such as demographic information, laboratory tests, medication usage, nursing records, and clinical notes. The prediction task focuses on hospital mortality, with the label defined as a binary variable, where 1 indicates death and 0 indicates survival. The dataset encompasses three primary data types: structured features, such as age, gender, admission type, and source; time-series features, such as daily counts of lab tests, medication orders, and caregiver activities; and text features, including physician notes and examination reports. These features collectively reflect patients' conditions and treatment histories, providing a robust foundation for multi-modal data modeling.

To ensure data quality and consistency, we applied several preprocessing steps. Missing values in structured data were handled using median imputation for numerical features and mode imputation for categorical features to minimize data loss while maintaining consistency across modalities. Numerical variables were scaled using min-max normalization to map values into the range [0, 1], ensuring comparability across features. Clinical notes were preprocessed through lowercasing, tokenization, and removal of stopwords to prepare the data for embedding extraction using the ClinicalBERT model. To avoid data leakage, features containing future information, such as length of stay, were excluded from the training data. This ensures that the model does not access information that would only be available after prediction time.

The dataset was divided into 70% for training, 15% for validation, and 15% for testing to ensure unbiased evaluation. Stratified sampling was applied to preserve the distribution of the mortality label

across subsets. Given the dataset's 12% mortality rate, class-weighted loss functions were implemented during training to address class imbalance and improve predictive performance.

4.2 Baseline Models and Experimental Setup

The experiment employed various traditional machine learning and deep learning models as baselines for comparative analysis, including random forests and logistic regression as traditional models. Random forests were used to model nonlinear interactions between features, while logistic regression served as a simple linear model to facilitate interpretability. For deep learning models, multilayer perceptrons (MLP), long short-term memory networks (LSTM), gated recurrent units (GRU), temporal convolutional networks (TCN), and Transformers were selected. These models are suitable for handling high-dimensional feature interactions, time-series modeling, and capturing complex sequential dependencies, enabling a comprehensive evaluation of different modeling approaches for the task. Table 1 shows the baseline model configurations. The dataset was split into 70% for training, 15% for validation, and 15% for testing. All models were fine-tuned using grid search or random search for hyperparameter optimization and evaluated through cross-validation. Evaluation metrics included Accuracy, F1 Score, and ROC-AUC to assess classification accuracy, performance under class imbalance, and overall discrimination capability. Additionally, confusion matrices and ROC curves were visualized to analyze classification performance and decision boundaries. The experiments were implemented using Python with TensorFlow frameworks and conducted under GPU-accelerated conditions for model training and testing.

Table 1: Baseline Model Configurations

Model	**Input Type**	**Key Features**	**Hyperparameters**
Logistic Regression	Structured data	Linear classification, interpretability	L2, Solver=lbfgs
Random Forest	Structured data	Nonlinear modeling, ensemble learning	n_estimators=100, max_depth=10
MLP	Structured data	High-dimensional feature interaction modeling	Hidden layers=2, Neurons=[128, 64], Dropout=0.5
LSTM	Time-series data	Sequential dependency modeling, long-term memory	Hidden size=128, Layers=2, Dropout=0.3
GRU	Time-series data	Simplified sequential modeling, efficient memory use	Hidden size=128, Layers=2, Dropout=0.3
TCN	Time-series data	Parallel temporal modeling, convolutional structure	Filters=64, Kernel size=3, Dropout=0.3
Transformer	Time-series & Text data	Attention mechanism, global context modeling	d_model=128, Heads=8, Layers=4, Dropout=0.1

4.3 Comparative Experiments

To evaluate the effectiveness of the proposed MADMN model, comparative experiments were conducted against various baseline models, including both traditional machine learning and deep learning approaches. The baseline models include logistic regression, random forest, multilayer perceptron (MLP), long short-term memory networks (LSTM), gated recurrent units (GRU), temporal convolutional networks (TCN), and Transformer. All models underwent the same preprocessing and data splitting procedures to ensure fair comparisons. Hyperparameters were tuned using grid search or random search, and the final performance was evaluated on the test set using Accuracy, F1 Score, and ROC-AUC as metrics. Additionally, ROC curves and confusion matrices were used to analyze classification performance and investigate error patterns.

Table 2 shows the performance comparison of models on the test set. The experimental results demonstrate that the MADMN model outperforms all baseline models across all evaluation metrics. On the test set, MADMN achieved an Accuracy of 0.882, an F1 Score of 0.765, and a ROC-AUC of

0.910. Compared to baseline models, traditional machine learning models showed reasonable performance with structured features but struggled to model complex interactions and temporal dynamics, as evidenced by the ROC-AUC of 0.812 for logistic regression and 0.842 for random forest. Deep learning models performed better, with the MLP achieving a ROC-AUC of 0.855, while LSTM and GRU leveraged sequential modeling capabilities to achieve 0.865 and 0.868, respectively. TCN and Transformer models further improved results, with ROC-AUC values of 0.872 and 0.878, but still fell short of the MADMN model. The p-values for all comparisons were less than 0.01, confirming that the observed improvements are statistically significant and unlikely to be due to random variation. This analysis demonstrates that MADMN's superior performance is attributable to its architecture and ability to capture complex multi-modal relationships rather than chance.

Table 2: Performance Comparison of Models on the Test Set

Model	**Accuracy**	**F1 Score**	**ROC-AUC**	**MCC**	**p-value (vs MADMN)**
Logistic Regression (LR)	0.78	0.65	0.812	0.452	< 0.001
Random Forest (RF)	0.81	0.685	0.842	0.498	< 0.001
Multilayer Perceptron (MLP)	0.825	0.7	0.855	0.525	< 0.001
LSTM	0.835	0.715	0.865	0.538	< 0.001
GRU	0.838	0.72	0.868	0.543	< 0.001
TCN	0.842	0.725	0.872	0.549	< 0.001
Transformer	0.85	0.73	0.878	0.557	< 0.001
MADMN (Proposed)	0.882	0.765	0.910	0.601	-

The superior performance of MADMN can be attributed to its cross-modal attention mechanism and dynamic memory network. The cross-modal attention mechanism effectively integrates structured, time-series, and textual features, leveraging complementary information across modalities. The dynamic memory network captures long-term dependencies in time-series data and highlights key features through attention mechanisms, enabling the model to better handle complex and evolving patient data. Figure 2 shows the ROC curves of different models. The ROC curves show that MADMN maintains a high true positive rate while keeping the false positive rate low, further confirming its superior classification capability and practical value.

In summary, the comparative experiments validate the advantages of the MADMN model in multimodal and dynamic modeling tasks. It outperforms baseline models in classification performance, robustness, and feature fusion, providing a more accurate and efficient solution for clinical risk prediction tasks.

4.4 Ablation Study

To further investigate the contributions of each component in the proposed MADMN model, an ablation study was conducted. The study analyzed the impact of removing or replacing key modules, including the cross-modal attention mechanism and the dynamic memory network, as well as the effects of excluding specific data modalities. The goal was to quantify the importance of these components and validate their contributions to model performance.

Table 3 shows results of ablation study and Figure 3 shows the ROC curves of ablation study. The first experiment removed the cross-modal attention mechanism and directly concatenated features from all modalities. The results showed a decline in F1 Score from 0.765 to 0.721 and ROC-AUC from 0.910 to 0.882, indicating that the attention mechanism effectively captured relationships between modalities and emphasized critical features. In the second experiment, the dynamic memory network was removed and replaced with a simple average pooling mechanism for temporal modeling. This change further reduced performance, with F1 Score dropping to 0.705 and ROC-AUC falling to 0.870, demonstrating that the memory network played a crucial role in capturing long-term dependencies and dynamic patterns.

To better assess the dynamic memory network, it was also replaced with LSTM and GRU variants. While these replacements partially recovered performance, achieving ROC-AUC values of 0.875 and

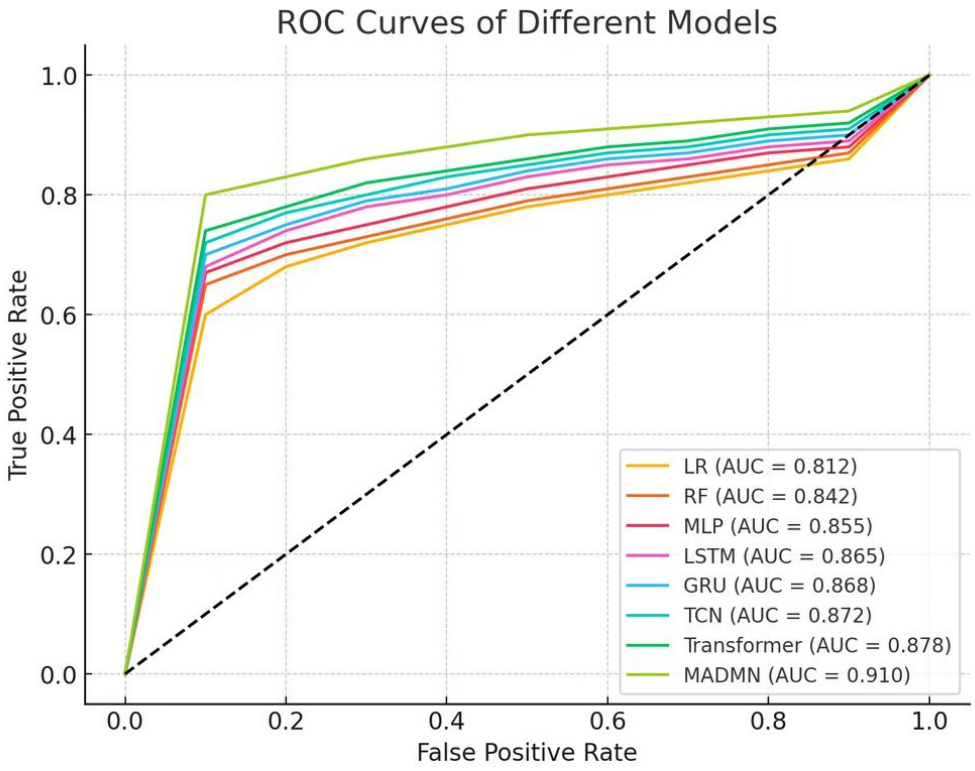


Figure 2: ROC Curves of Different Models

Table 3: Results of Ablation Study

Model Variant	**Accuracy**	**F1 Score**	**ROC-AUC**
Full Model (MADMN)	0.882	0.765	0.910
No Attention	0.850	0.721	0.882
No Memory Network	0.832	0.705	0.870
Memory as LSTM	0.842	0.715	0.875
Memory as GRU	0.845	0.720	0.878
No Structured Data	0.841	0.728	0.873
No Time-Series Data	0.835	0.713	0.868
No Text Data	0.843	0.732	0.875

0.878 respectively, they still underperformed compared to the full MADMN model. This highlights the effectiveness of the proposed memory design for modeling temporal dependencies.

Further experiments examined the importance of different data modalities by removing structured features, time-series features, and text features. The results showed noticeable declines in performance when any modality was excluded, with F1 Scores dropping to 0.728, 0.713, and 0.732, respectively. Among these, removing time-series data resulted in the most significant drop, underscoring its importance for capturing temporal dynamics in patient conditions.

Finally, alternative designs for the attention mechanism were tested by replacing it with simple weighted sums and average pooling. Both modifications resulted in lower performance, with ROC-AUC values of 0.879 and 0.870, respectively, indicating that the cross-modal attention mechanism effectively captured interactions between different modalities and enhanced feature representation.

The ablation study results, summarized in the table below, clearly demonstrate that the cross-modal attention mechanism and dynamic memory network are critical for achieving optimal performance. The MADMN model leverages these components to effectively integrate multi-modal data and model temporal dependencies, providing superior classification accuracy and robustness compared to simplified designs or excluded features.

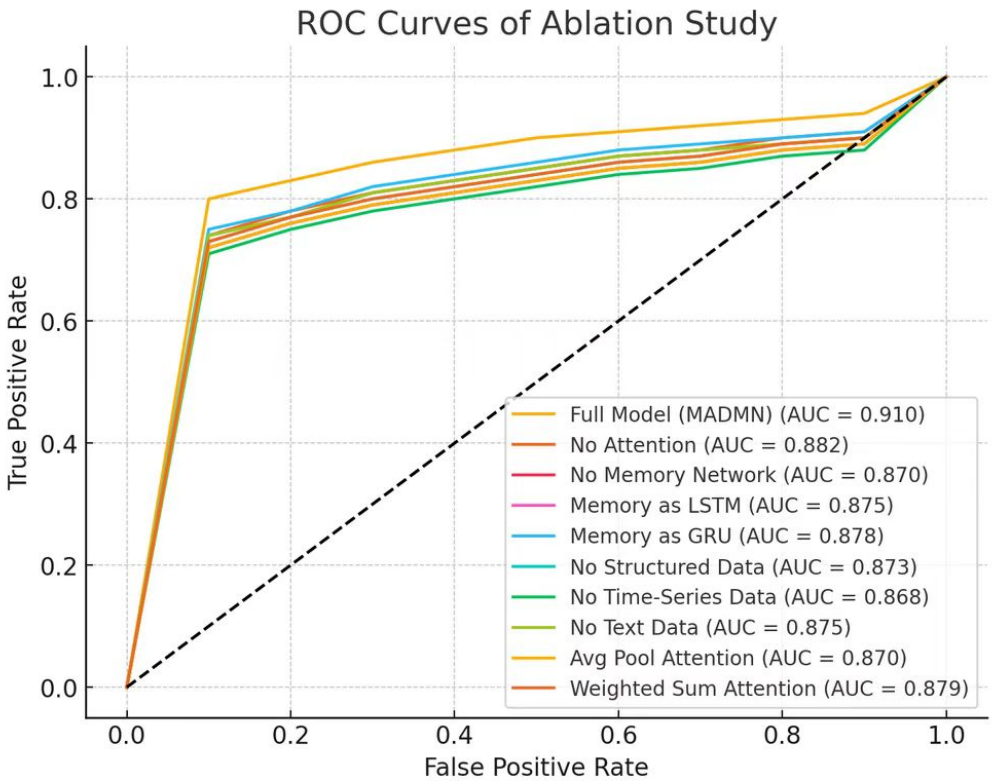


Figure 3: ROC Curves of Ablation Study

5 Discussions

5.1 Theoretical Implications

Through the SHAP analysis of the MADMN model, several important theoretical implications emerged. First, the analysis revealed that the model is able to identify clinically significant features for mortality risk prediction. For instance, age is a critical factor, with older patients showing higher predicted mortality risks, especially when age exceeds a certain threshold, where the risk rises sharply. Length of stay also plays a crucial role, as longer hospitalizations often indicate more severe conditions and greater treatment needs, thus elevating the mortality risk. Admission type and admission pathways are also significant structural features, with patients admitted through emergency or referral pathways tending to have higher risks due to the severity of their conditions. These findings suggest that future theoretical research could continue to focus on the importance of features in combination with clinical context, thereby improving model accuracy and reliability.

Moreover, the analysis revealed interactions between static and dynamic features, with dynamic features proving particularly important when patients' conditions change over time. The model successfully captured changes in dynamic features such as medication frequency, vital sign fluctuations, and lab test variations, all of which were strongly associated with increased mortality risk. For example, frequent adjustments in medication dosages or changes in lab tests often indicate clinical deterioration, which raises mortality risk. This finding provides new perspectives for theoretical research, emphasizing the importance of joint modeling of static and dynamic features in predictive models to improve prediction accuracy.

In addition, the study showed that the relationship between medication frequency and mortality risk is nonlinear. Initially, frequent medication use may decrease risk because it suggests active treatment. However, excessive intervention beyond a certain threshold could signal worsening conditions, leading to a sharp increase in risk. This phenomenon confirms that the model effectively captures nonlinear relationships and threshold effects, providing valuable insights into the modeling of complex clinical data.

Finally, counterfactual analysis demonstrated the impact of various interventions on mortality

risk prediction. For instance, reducing medication frequency or adjusting lab test intervals after stabilization led to a noticeable decrease in predicted mortality risk. This suggests that counterfactual analysis not only has practical value but also provides a new direction for the theoretical development of predictive models, especially in simulating the effects of different treatment interventions.

5.2 Practical Implications

From a practical perspective, the MADMN model offers several key insights for clinical practice, particularly in treatment optimization, risk assessment, and personalized medical decision-making. First, the MADMN model enables precise mortality risk prediction for individual patients, providing tailored treatment recommendations. For high-risk patients, reducing medication frequency or adjusting lab test frequencies can significantly lower predicted mortality risk. Conversely, for medium-risk patients, increased care interventions and more frequent monitoring might be required to detect potential health issues earlier.

Furthermore, counterfactual analysis supports clinical decision-making by simulating the effects of different treatment strategies. Doctors can model different interventions, such as altering medication dosages or modifying admission pathways, to assess their impact on mortality risk. This enables the identification of the most effective treatment strategies and supports evidence-based decisions. Such analyses help doctors understand the potential effects of treatment plans and provide more personalized care for patients.

In terms of resource allocation, the MADMN model can assist hospitals in optimizing their resource distribution. By assessing patient risk levels, hospitals can prioritize high-risk patients for intensive monitoring and care, ensuring that medical resources are used effectively. This not only improves hospital operational efficiency but also ensures that patients receive the best treatment according to their condition.

Moreover, when integrated into hospital information systems, the MADMN model can provide real-time mortality risk assessments and dynamic monitoring. By continuously tracking patient data such as vital signs, medication usage, and lab test results, the system can promptly identify high-risk patients and alert clinical staff, allowing for timely interventions. This helps hospitals respond quickly to changes in patient conditions and ensure timely treatment.

Finally, the interpretability of the model is a major advantage in practical applications. The SHAP analysis provides transparency, allowing clinicians to understand the reasoning behind the model's predictions. This not only increases trust in the AI model but also facilitates its widespread adoption in clinical settings. As AI technology continues to develop, enhancing model interpretability will be key to its broader application in healthcare.

6 Conclusions

This paper proposes MADMN, a Multi-modal Attention and Dynamic Memory Network, for early mortality risk prediction based on electronic medical records. The model addresses the challenges posed by the heterogeneous and dynamic nature of EMR data by integrating multi-modal feature extraction, cross-modal attention mechanisms, and dynamic memory networks. It effectively models structured, time-series, and text data, capturing both long-term dependencies and short-term variations in patient conditions.

Experimental results demonstrate that MADMN significantly outperforms traditional machine learning and deep learning baseline models in terms of accuracy, F1 score, and ROC-AUC. This validates its ability to process multi-modal feature interactions and dynamic changes effectively. The ablation study further confirms the critical contributions of the cross-modal attention mechanism and dynamic memory network to model performance and interpretability. SHAP analysis verifies the interpretability of the model by highlighting key features such as age, length of stay, and dynamic clinical interactions. These findings align closely with clinical insights, ensuring transparency and trust in the model's predictions. Additionally, counterfactual analysis showcases MADMN's potential in personalized treatment planning by evaluating the impact of hypothetical changes in patient conditions,

thus supporting clinical interventions and resource optimization. In conclusion, the proposed multimodal risk prediction model has the potential to significantly enhance the application of EMRs in clinical decision-making and risk assessment. By leveraging complementary information from heterogeneous data sources, dynamically modeling temporal patterns, and providing interpretable outputs, the model contributes to advancing personalized and precision medicine, offering new solutions for improving patient care.

Despite its promising performance, this study has several limitations. First, the model relies on the MIMIC-III dataset, which primarily represents ICU patients from a specific region in the United States. This may limit its generalizability to other populations or healthcare systems. Future work could focus on testing the model on larger and more diverse datasets to enhance its applicability. Second, the text data processing relies heavily on the pre-trained ClinicalBERT model, which may not fully capture semantic information in non-English texts or varying medical terminologies. Future research could explore multilingual and cross-cultural text modeling approaches to improve generalizability.

The MADMN model demonstrates strong performance in early mortality risk prediction; however, several limitations must be acknowledged. First, the model was trained and validated using the MIMIC-III dataset, which primarily represents ICU patients in the United States. This may limit its generalizability to other geographic regions and healthcare systems. Future work should test the model on more diverse datasets and explore domain adaptation techniques to improve adaptability across populations. Second, the model's complexity may pose challenges for deployment in resource-constrained environments. Developing lightweight versions of MADMN using model compression or knowledge distillation could address this issue and enable real-time applications. Third, the reliance on ClinicalBERT for text processing may restrict performance in non-English contexts. Incorporating multilingual models and data augmentation methods could enhance the model's versatility. Finally, while the model improves interpretability through SHAP analysis and counterfactual simulations, further enhancements using attention visualizations and causal inference methods could provide deeper insights into feature interactions and temporal patterns. Future extensions could also focus on broader clinical applications, such as readmission prediction, length-of-stay estimation, and complication detection, to increase the model's utility in clinical practice.

References

[1] Banerjee, J., Taroni, J. N., Allaway, R. J., Prasad, D. V., Guinney, J., & Greene, C. (2023). Machine learning in rare disease. *Nature Methods*, 20(6), 803-814.

[2] Rahman, S., Hasan, M., & Sarkar, A. K. (2023). Prediction of brain stroke using machine learning algorithms and deep neural network techniques. *European Journal of Electrical Engineering and Computer Science*, 7(1), 23-30.

[3] Oh, J.W. (2022). A study on digital healthcare service in big data environment: focusing on diagnosis of hyperlipidemia based on diagnostic testing. *Journal of System and Management Sciences*, 12(3), 345-360. https://doi.org/10.33168/JSMS.2022.0317

[4] Haque, R., Ho, S.B., Chai, I., Abdullah, A. (2023). Improved Adam-based feedforward deep neural network model for personalized asthma predictions. *Journal of System and Management Sciences*, 13(2), 241-257. https://doi.org/10.33168/JSMS.2023.0217

[5] Kim, Y.S., Song, H.J., Han, J. (2022). A Deepfake-Based Deep Learning Algorithm for Medical Data Manipulation Detection. *Journal of System and Management Sciences*, 12(1), 13-24. https://doi.org/10.33168/JSMS.2022.0102

[6] Yuantai, X. (2023). Discrete Chaotic Fuzzy Neural Network (DC-FNN) Based Smart Watch Embedded Devices for Sports and Health Monitoring. *Tehnički vjesnik*, 30(6), 1784-1790. https://doi.org/10.17559/TV-20230222000374

[7] Aminizadeh, S., Heidari, A., Toumaj, S., Darbandi, M., Navimipour, N. J., Rezaei, M., ... & Unal, M. (2023). The applications of machine learning techniques in medical data processing

based on distributed computing and the Internet of Things. *Computer Methods and Programs in Biomedicine*, 107745.

[8] Feng, J., Zhang, R., Chen, D., Shi, L., & Li, Z. (2024). Automated generation of ICD-11 cluster codes for precision medical record classification. *International Journal of Computers Communications & Control*, 19(1). https://doi.org/10.15837/ijccc.2024.1.6251

[9] Morid, M. A., Sheng, O. R. L., & Dunbar, J. (2023). Time series prediction using deep learning methods in healthcare. *ACM Transactions on Management Information Systems*, 14(1), 1-29.

[10] Foersch, S., Glasner, C., Woerl, A. C., Eckstein, M., Wagner, D. C., Schulz, S., ... & Jesinghaus, M. (2023). Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nature Medicine*, 29(2), 430-439.

[11] Fremond, S., Andani, S., Wolf, J. B., Dijkstra, J., Melsbach, S., Jobsen, J. J., ... & Bosse, T. (2023). Interpretable deep learning model to predict the molecular classification of endometrial cancer. *The Lancet Digital Health*, 5(2), e71-e82.

[12] Saeedi, S., Rezayi, S., Keshavarz, H., & R. Niakan Kalhori, S. (2023). MRI-based brain tumor detection using convolutional deep learning methods. *BMC Medical Informatics and Decision Making*, 23(1), 16.

[13] Curila, S., Buciu, I., Grava, C., Trip, D. N., & Straciuc, O. M. (2024). COVID-19 lung infection segmentation using statistics and active contour. *International Journal of Computers Communications & Control*, 19(6). https://doi.org/10.15837/ijccc.2024.6.6862

[14] Mahmood, M. A., Alsalem, K., Elbashir, M., El-Ghany, S. A., & Abd El-Aziz, A. (2024). Acute knee injury detection with MRI. *International Journal of Computers Communications & Control*, 19(5). https://doi.org/10.15837/ijccc.2024.5.6648

[15] Vrbančič, G., Pečnik, Š., & Podgorelec, V. (2022). Hyper-parameter Optimization for COVID-19 X-ray Classification. *Computer Science and Information Systems*, 19(1), 327-352. https://doi.org/10.2298/CSIS210209056V

[16] Singh, A., Pannu, H. S., & Malhi, A. (2022). Explainable Information Retrieval using Deep Learning for Medical Images. *Computer Science and Information Systems*, 19(1), 277-307. https://doi.org/10.2298/CSIS201030049S

[17] Babu, P. A., Rao, B. V. S., Reddy, Y. V. B., Kumar, G. R., Rao, J. N., Reddy, S. K., & Kumar, G. S. (2023). Optimized CNN-based brain tumor segmentation and classification using artificial bee colony and thresholding. *International Journal of Computers Communications & Control*, 18(1). https://doi.org/10.15837/ijccc.2023.1.4577

[18] Lee, M. H., Zea, R., Garrett, J. W., Summers, R. M., & Pickhardt, P. J. (2025). AI-based abdominal CT measurements of orthotopic and ectopic fat predict mortality and cardiometabolic disease risk in adults. *European Radiology*, 35(1), 520-531.

[19] Placido, D., Yuan, B., ... & Sander, C. (2023). A deep learning algorithm to predict pancreatic cancer. *Nature Medicine*, 29(5), 1113-1122.

[20] Forrest, I. S., Petrazzini, B. O., ... & Do, R. (2023). Marker for coronary artery disease validation. *The Lancet*, 401(10372), 215-225.

[21] Renc, P., Jia, Y., Samir, A. E., Was, J., Li, Q., Bates, D. W., & Sitek, A. (2024). Zero shot health trajectory prediction using transformer. *NPJ Digital Medicine*, 7(1), 256.

[22] Yang, J., Soltan, A. A., ... & Clifton, D. A. (2023). Bias mitigation for clinical ML. *Nature Machine Intelligence*, 5(8), 884-894.

[23] Pfob, A., Mehrara, B. J., ... & Sidey-Gibbons, C. (2023). Predicting breast cancer surgery outcomes. *Annals of Surgery*, 277(1), e144-e152.

[24] Yashudas, A., Gupta, D., Prashant, G. C., Dua, A., AlQahtani, D., & Reddy, A. S. K. (2024). DEEP-CARDIO: Recommendation System for Cardiovascular Disease Prediction using IOT Network. *IEEE Sensors Journal*.

[25] Wan, X., & Tian, L. (2024). User stress detection using social media text: A novel machine learning approach. *International Journal of Computers Communications & Control*, 19(5). https://doi.org/10.15837/ijccc.2024.5.6772

[26] Zhu, K., Li, R., Yao, P., Yu, H., Pan, A., Manson, J. E., ... & Liu, G. (2025). Proteomic signatures of healthy dietary patterns are associated with lower risks of major chronic diseases and mortality. *Nature Food*, 6(1), 47-57.

[27] Wang, Q., Chen, G., Jin, X., Ren, S., Wang, G., Cao, L., & Xia, Y. (2023). BiT-MAC: Mortality prediction by bidirectional time and multi-feature attention coupled network on multivariate irregular time series. *Computers in Biology and Medicine*, 160, 106586. https://doi.org/10.1016/j.compbiomed.2023.106586

[28] Guo, C., Liu, M., & Lu, M. (2021). A dynamic ensemble learning algorithm based on K-means for ICU mortality prediction. *Applied Soft Computing*, 107, 107166. https://doi.org/10.1016/j.asoc.2021.107166

[29] Liu, M., Guo, C., & Guo, S. (2023). An explainable knowledge distillation method with XGBoost for ICU mortality prediction. *Computers in Biology and Medicine*, 157, 106466. https://doi.org/10.1016/j.compbiomed.2023.106466

[30] Ren, S., Guo, B., Li, K., Wang, Q., Yu, Z., & Cao, L. (2022). Coupledmuts: Coupled multivariate utility time-series representation and prediction. *IEEE Internet of Things Journal*, 9(14), 11520–11530. https://doi.org/10.1109/JIOT.2022.3143615

[31] Shi, Z., Wang, S., Yue, L., Zhang, Y., Adhikari, B. K., Xue, S., Zuo, W., & Li, X. (2023). Dual-core mutual learning between scoring systems and clinical features for ICU mortality prediction. *Information Sciences*, 627, 119042. https://doi.org/10.1016/j.ins.2023.119042

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

Cite this paper as:

Chen, Y.; Chen, K. (2025). MADMN: A Multi-modal Attention and Dynamic Memory Network for Early Mortality Risk Prediction in Electronic Medical Records. *International Journal of Computers Communications & Control*, 20(3), 6968, 2025. https://doi.org/10.15837/ijccc.2025.3.6968