Information Fusion 115 (2025) 102780

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors

Jianhui Lv [a,b,*], Byung-Gyu Kim [c], B.D. Parameshachari [d], Adam Slowik [e], Keqin Li [f]

[a] The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China
[b] Peng Cheng Laboratory, Shenzhen, China
[c] Sookmyung Women's University, Seoul, Republic of Korea
[d] Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka 560064, India
[e] Koszalin University of Technology, Koszalin, Poland
[f] State University of New York, New Paltz, NY 12561, USA

ARTICLE INFO

Keywords:
Large model
Hyperscale healthcare data fusion
Complex multi-sensors
Residual connection
Dynamic ReLU

ABSTRACT

In the era of big data and artificial intelligence, healthcare data fusion analysis has become difficult because of the large amounts and different types of sources involved. Traditional methods are ineffective at processing and examination procedures for such complex multi-sensors of hyperscale healthcare data. To address this issue, we propose a novel large model-driven approach for hyperscale healthcare data fusion analysis in complex multi-sensor multi-sensors. Our method integrates data from various medical sensors and sources, using large models to extract and fuse information from structured and unstructured healthcare data. Then, we integrate these features with structured data using a hierarchical residual connected LSTM network, enhancing the model's ability to capture local and global context. Furthermore, we introduce a dynamic ReLU activation function and attention mechanism that allow us to adjust the depth of our networks dynamically while focusing only on relevant information. The experiments on MIMIC-III and eICU-CRD datasets demonstrate the superiority of the proposed method in terms of accuracy, efficiency, and robustness compared to state-of-the-art methods. Therefore, the proposed method provides valuable insights into the potential of large model-driven approaches for tackling the challenges of hyperscale healthcare data fusion analysis in complex multi-sensors.

1. Introduction

Healthcare data has rapidly grown in the past few years since electronic health records, wearable device adoption, and advanced medical imaging technologies [1]. All this provides an opportunity for healthcare research based on data, as well as applications like never before. However, traditional methods of analyzing information face significant challenges because of the amount, types, and speed at which they are produced within the hyperscale healthcare data [2]. The complexity and unstructured nature of the healthcare data, high dimensionality associated with genomic data, or even heterogeneity among different sources that make up this wide-ranging landscape make it difficult to extract useful insights from such a wealth of knowledge about healthcare [3]. Therefore, researchers have started using machine learning models like those used in deep learning architectures or transformer-based systems to solve these problems since they have proved very successful in working with complex and high-dimensional datasets across various fields [4,5]. Such large models can detect fine-grained details hidden deep down in massive datasets, thereby enabling us to leverage them so doctors can get accurate, personalized solutions for diagnosis treating monitoring people's health based on their previous history, etcetera driven by large amounts of individualized information [6,7]. The complexity of healthcare data has further increased with integrating multi-sensor systems in medical environments, producing diverse data streams that need to be fused and analyzed collectively. This multi-sensor fusion adds another layer of complexity to the already challenging task of healthcare data analysis.

In order to deal with the problems of hyperscale healthcare data fusion analysis in complicated settings, researchers have considered many machine learning methods and deep learning techniques. Among

* Corresponding author.
E-mail addresses: lvjh@pcl.ac.cn (J. Lv), bg.kim@sookmyung.ac.kr (B.-G. Kim), paramesh@nmit.ac.in (B.D. Parameshachari), adam.slowik@tu.koszalin.pl (A. Slowik), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.inffus.2024.102780
Received 9 July 2024; Received in revised form 4 October 2024; Accepted 31 October 2024
Available online 4 November 2024

them are large models like BERT or GPT-3 that best handle unstructured medical text complexity [8,9,10]. These models were taught on large amounts of text data. They showed impressive results in different natural language processing tasks such as named entity recognition, relation extraction, or question answering [11,12,13]. Based on self-supervised learning and transformer, these large models can grasp deep semantic sense from unstructured texts and create well-contextualized representations that would serve as a good basis for further downstream task improvement. Also, they allow us to understand complicated relations between words in sentences, which is only possible for small models with transformers. Moreover, their ability to detect complex patterns from massive amounts of information makes them perfect for working with various healthcare knowledge. However, applying this model to a healthcare environment with complex conditions is difficult because we need high accuracy rates and interpretability.

Even though they can be very helpful in analyzing healthcare data, large models are only partially applicable to complex multi-sensors since healthcare data is domain-specific due to many expert terms, acronyms, and writing styles that differ significantly from those used in training such models on general language [14]. The multiple sensors include a wide range of devices such as physiological monitors (ECG, blood pressure, oxygen saturation, respiratory rate), imaging devices (X-ray, CT, MRI, ultrasound), laboratory equipment (blood analyzers, urine analyzers), wearable devices (continuous glucose monitors, activity trackers), and environmental sensors (room temperature, humidity, air quality). The multi-sensor approach allows the model to comprehensively view a patient's condition by integrating diverse data streams. Each sensor type contributes unique information processed and fused by the model to improve the accuracy and robustness of healthcare predictions and analyses. This multi-modal data integration enables LM-HHDA to understand patient health status and potential outcomes better. The following issue arises from this discrepancy between domains: How can these large models work well for healthcare? Some acceptable methods should be developed, like fine-tuning with medical corpora, specific knowledge through entity embedding, or integrating knowledge graphs based on domains. For instance, fusing electronic health records with real-time sensor data from wearable devices can enable early detection of deteriorating conditions in chronic disease patients.

The other problem is integrating unstructured data features extracted by large models with structured data features such as numeric measurements or categorical variables commonly found in healthcare datasets [15]. These data types are so diverse that designing a single architecture that considers both types simultaneously (continuous vs discrete) takes work. For example, whether height is continuous, whereas ethnicity could be considered discrete. Therefore, any model that handles all inputs should be able to process mixed-type inputs well – this means effectively combining information from different sources within one framework. One possible solution might involve creating new hybrid models that use parts of traditional machine learning systems like decision trees and random forests [16]. In order to have a broader coverage over different input patterns in healthcare situations, such models use the power of large models.

Moreover, there is a computational aspect because healthcare datasets can be extremely large-scale, sometimes involving millions upon millions of patient records along with terabytes of imaging or genomic data [17]. In other words, imagine having billions of rows each day where every row corresponds to another image file! Such amounts pose serious challenges when training and during the inference phase, i.e., making predictions given an already trained model. Therefore, any such system must consider space requirements (memory) and time efficiency (speed) when processing such volumes. Otherwise, training may take years instead of minutes or vice versa. Several methods could be used for this purpose: distributed computing, parallel processing, etc., but the most important is achieving a balance between memory usage and speed.

To conquer these hurdles, diverse disciplines such as natural language processing, machine learning, data engineering, and healthcare informatics need to be brought together [18,19,20]. Researchers can unleash the full potential of large models if they create inventive approaches that solve peculiar problems while analyzing healthcare-related information within complicated settings; this will transform diagnosis methods, treatment procedures, and preventive measures against diseases.

The main contributions of this paper are the following:

- We use large models for semantic feature extraction in unstructured medical text and utilize them alongside structured data features through a hierarchical residual connected LSTM network.
- We propose a dynamic ReLU activation function that can adjust the depths of the network dynamically depending on input data to allow the model to capture more complicated patterns without compromising computational efficiency.
- We adopt an attention mechanism focusing on relevant information to increase accuracy and interpretability while suppressing noise.

The remainder of this paper is organized as follows. Section II presents the proposed methodology. Section III presents the experimental results and analysis. Finally, Section IV concludes the paper.

2. Methodology

The proposed LM-HHDA method is designed to handle the complexities of multi-sensor healthcare environments by fusing data from various sources such as wearable devices, medical imaging equipment, and electronic health records. The method employs advanced fusion techniques at different levels of abstraction to ensure a comprehensive analysis of the heterogeneous data streams. Fig. 1 displays the general structure of the proposed large model-driven hyperscale healthcare data fusion analysis method, consisting of five main parts:

Text Representation: The text representation component uses large pre-trained models to extract rich semantic features from unstructured medical text. These models are fine-tuned on domain-specific healthcare corpora to capture the nuances and complexities of medical language. By encoding the vast amounts of healthcare data using these powerful language models, the proposed framework can effectively capture the intricate relationships and patterns within the text data.

Hierarchical Residual Connected LSTM: The hierarchical residual connected LSTM component is designed to capture long-term dependencies and hierarchical structures in the hyperscale healthcare data. Integrating the encoded text features with structured data features enables the proposed framework to effectively model the complex relationships and patterns within the healthcare data. The residual connections in the LSTM architecture alleviate the vanishing gradient problem and allow for the efficient flow of information through the network. The hierarchical structure enables the framework to learn meaningful representations at different levels of granularity, from individual patient records to higher-level abstractions.

Local Feature Modeling: The multi-scale convolutional architecture component is designed to extract local features from the input data at different scales. This component captures the spatial and temporal patterns in the healthcare data by employing convolutional neural networks (CNNs) with varying kernel sizes and dilation rates. The multi-scale approach enables the proposed framework to learn hierarchical representations of the local features, from fine-grained details to higher-level abstractions. This component complements the hierarchical residual connected to LSTM by capturing the local context and patterns, which is crucial for understanding the complex relationships and dependencies within the healthcare data.

Dynamic ReLU Activation: The dynamic ReLU activation component introduces an adaptive activation function that adjusts its parameters

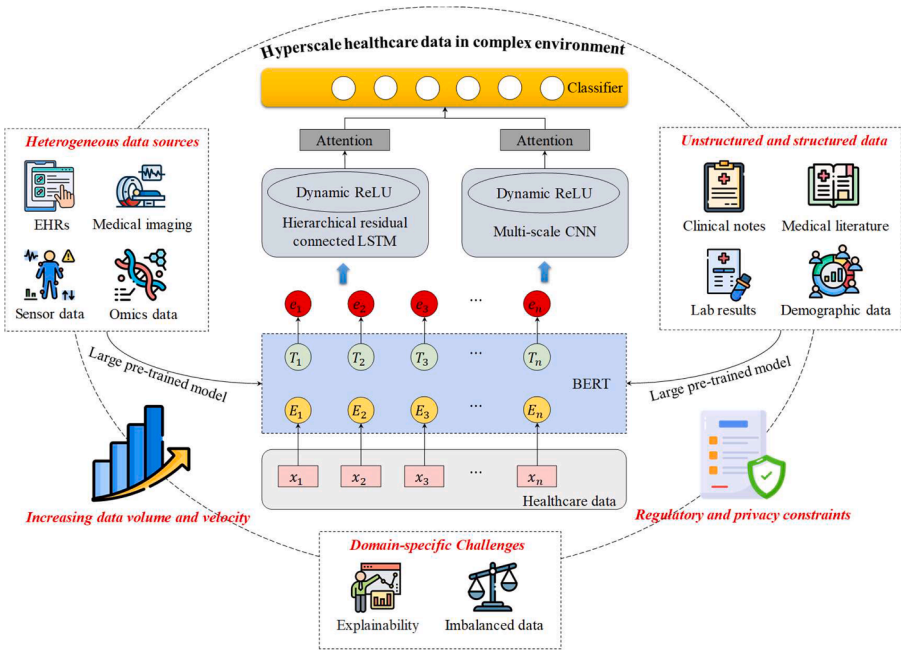

Fig. 1. Overall framework.

based on the input data. By dynamically adapting to healthcare data's complex and heterogeneous nature, this component enhances the proposed framework's ability to capture intricate patterns and relationships. The dynamic ReLU activation function allows the model to learn more expressive and flexible representations, which is crucial for handling the diverse and evolving healthcare data landscape. By adapting to the specific characteristics of the input data, the dynamic ReLU activation component enables the proposed framework to capture better the underlying structure and semantics of the healthcare data. This adaptive activation function also helps improve the model's robustness and generalization ability, making it more suitable for real-world healthcare applications.

Relationship Prediction: Finally, an attention mechanism is employed that dynamically concentrates on key features required for predicting relationships between objects/entities within large models together with hierarchy residuals connected LSTMs while at the same time suppressing noise as well as irrelevant information.

In the choice of its specific components, the proposed method does this to capture complex patterns and dependencies in healthcare data. The hierarchical residual connected LSTM is used to model the sequential nature of healthcare data while dealing with the vanishing gradient problem. The hierarchical residual connected LSTM allows it to study more expressive representations by enabling data flow through multiple layers. Multi-scale convolutional architecture, on the other hand, is used to capture local patterns and extract discriminative features at different granularities. Combining these components enables the proposed method to effectively learn global and local patterns in healthcare data, leading to improved performance in various analysis tasks.

Also, LM-HHDA must include a missing data imputation module based on matrix factorization techniques. This module estimates missing values by leveraging correlations between observed data points, ensuring robust performance even with incomplete datasets.

2.1. Text representation

In large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors, it is vital to represent and encode large amounts of unstructured medical text effectively. We employ BERT to harness rich semantic information contained within these texts. BERT has been trained on massive amounts of text data from different fields and have shown impressive results in many natural language processing tasks, such as relation extraction, named entity recognition, and answering questions about a passage [21,22,23]. In addition to unstructured medical text, our text representation component is adapted to handle textual data from multiple sensors, such as patient monitoring devices and lab equipment. This multi-sensor text fusion allows a more comprehensive understanding of the patient's condition.

To make these large models suitable for healthcare-specific applications, we start by fine-tuning them on large medical corpora like electronic health records, clinical notes, as well as medical literature. This process allows the models to adapt their understanding towards unique terms used within this domain and also get accustomed to various writing styles employed in medical texts, thus enabling them to capture more semantic information specific to that area.

For given input medical text sequence $X = [x_1, x_2, \ldots, x_n]$ where x_i represents i-th token in the sequence. We tokenize the text first and then get corresponding embeddings for each token $E = [e_1, e_2, \ldots, e_n]$ using the fine-tuned model:

$$E = \text{BERT}(X) = [e_1, e_2, \ldots, e_n], e_i \in \mathbb{R}^d. \tag{1}$$

where d is the dimensionality of the token embeddings.

Initially, a named entity recognition module is used to identify the medical entities within the input sequence. The module can either be a rule-based system or a pre-trained model. Let $M = [m_1, m_2, \ldots, m_k]$ denote the recognized medical entities where each m_i stands for the i-th entity mention.

Then, we calculate attention scores α_{ij} between each token x_i and each entity mention m_j in terms of an entity-aware bilinear attention function:

$$\alpha_{ij} = \text{softmax}\left(e_i^T W_a e_{m_j}\right). \tag{2}$$

where $W_a \in \mathbb{R}^{d \times d}$ is a learnable weight matrix, and e_{m_j} is the token embedding of the first token in the entity mention m_j.

We compute the entity-aware context representation c_i for each token x_i by taking a weighted sum of the token embeddings based on the attention scores:

$$c_i = \sum_{j=1}^{k} \alpha_{ij} e_{m_j}. \tag{3}$$

We concatenate the token embeddings e_i with their corresponding entity-aware context representations c_i to get the last text representation $T = [t_1, t_2, ..., t_n]$.

$$t_i = [e_i; c_i], t_i \in \mathbb{R}^{2d}. \tag{4}$$

where $[;]$ denotes the concatenation operation.

Moreover, we extend the text representation to hierarchical embeddings to represent medical text as objects that reflect its hierarchical structure.

For instance, let $S = [s_1, s_2, ..., s_m]$ denote the sequence of sentences in a medical document where each sentence s_i has token embeddings $E_i = [e_{i1}, e_{i2}, ..., e_{in_i}]$. We may calculate sentence embeddings s_i' through applying mean pooling over token embeddings:

$$s_i' = \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}. \tag{5}$$

Similarly, we can calculate the paragraph and document embeddings by using mean pooling or attention mechanisms on the sentence and paragraph embeddings, respectively.

When these hierarchical embeddings are integrated into text representation, a model may apprehend various relationships and dependencies in hyperscale medical text data, leading to more accurate healthcare data fusion analysis in complex multi-sensors.

LM-HHDA employs a two-stage missing data imputation strategy. In the first stage, matrix factorization techniques estimate missing values by leveraging correlations between observed data points. In the second stage, a deep learning-based imputation model refines these initial estimates by capturing complex non-linear relationships between variables. This hybrid approach ensures robust performance with incomplete datasets while maintaining computational efficiency.

The large model-driven approach that represents the text utilizes fine-tuned pre-trained models to capture the semantic information in medical texts. The entity-aware attention mechanism is introduced alongside hierarchical embeddings to enhance this part's understanding capability for complex relationships between different medical entities and between those entities themselves at different levels. This strong foundation created by robust, comprehensive text representation enables accurate, insightful healthcare data fusion analysis under complicated circumstances for subsequent components comprising the rest of the parts, including decision-making processes.

2.2. Residual connection

To get the most accurate and complete findings in large model-driven hyperscale healthcare data fusion analysis of complex multi-sensors, capturing long-term dependencies and hierarchical structures in the data is important. However, when faced with more complex datasets that require deep neural networks, training such a model becomes harder due to the vanishing gradient problem [24]. To address this issue, we introduce a hierarchical residual connected LSTM architecture. The main concept behind residual learning is adding shortcut connections that bypass one or more layers, thus enabling models to learn residual functions with reference to these layers' inputs. This has been proven useful in many cases for training deeper neural networks as well as improving their performance [25]. The hierarchical residual connected LSTM architecture is further enhanced to handle multi-sensor inputs. Each LSTM layer can process data from different sensors, with residual connections allowing for effective fusion of these diverse data streams.

In our suggested design, we include residual connections within LSTM recurrent cells so that they can effectively gather information about long-term dependencies across different levels of abstraction inherent in large scale healthcare datasets. For instance, Fig. 2 shows an example of modifying a standard LSTM cell by adding another fully connected layer.

In order to get hidden state representations $H = [h_1, h_2, ..., h_n]$ with $h_i \in \mathbb{R}^h$ (h being dimensionality of hidden state), this sequence should first pass through an LSTM layer.

To add residual connections, we apply a fully connected layer followed by a non-linear activation function on top of each transformed hidden state h_i.

$$\widetilde{h}_i = \mathrm{ReLU}(W_h h_i + b_h). \tag{6}$$

where $W_h \in \mathbb{R}^{h \times h}$ and $b_h \in \mathbb{R}^h$ are learnable parameters.

Then, we add a residual connection that bypasses the fully connected layer and the non-linear activation, letting the model learn the residual function.

$$h_i' = h_i + \widetilde{h}_i. \tag{7}$$

These residual connections have been introduced to simplify a model to learn the identity function. For a better representation of hierarchical structures in data by the model, we apply residual connected LSTM on multiple layers, as indicated by Fig. 3.

In this hierarchical architecture, we stack multiple LSTM layers connected through residuals, where each layer's output serves as the next layer's input. With such a design, the model can understand the input sequence in different levels of granularity—low-level attributes and high-level semantics.

Let L be the number of layers in the hierarchical residual network LSTM. The hidden state representation at layer l and time step i is given by:

$$h_i^{(l)} = \mathrm{LSTM}\left(h_i^{(l-1)}, h_{i-1}^{(l)}\right) + \widetilde{h}_i^{(l)}. \tag{8}$$

where $h_i^{(l-1)}$ is the hidden state from the previous layer at time step i, $h_{i-1}^{(l)}$ is the hidden state from the previous time step at the current layer, and $\widetilde{h}_i^{(l)}$ is the transformed hidden state obtained by fully connected layer and non-linear activation applied to $h_i^{(l)}$.

These hierarchical residual connections allow the model better to capture complex hierarchical structures in hyperscale healthcare data, thus enabling accurate and comprehensive data analysis within complicated multi-sensors.

Besides, we can also include more transformations in residual connections to handle the diversity and complexity of healthcare data. For instance, convolutional layers or self-attention mechanisms can be utilized as additional transformations. The example below demonstrates a 1D convolutional residual connection through the application of convolutions on hidden states:

$$\widetilde{h}_i^{(l)} = \mathrm{ReLU}\left(W_c * h_i^{(l)} + b_c\right). \tag{9}$$

where $W_c \in \mathbb{R}^{k \times h \times h}$ and $b_c \in \mathbb{R}^h$ are learnable parameters, k is the kernel size, and $*$ denotes the convolution operation.

Similarly, we introduce a self-attention residual connection by implementing a self-attention mechanism on the hidden states.

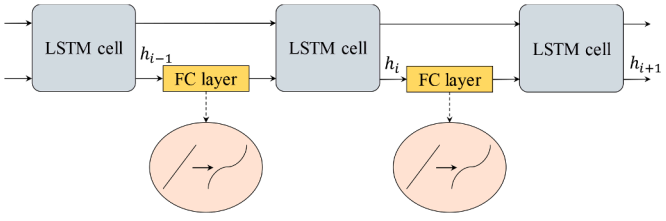

Fig. 2. LSTM cell with fully connected layer.

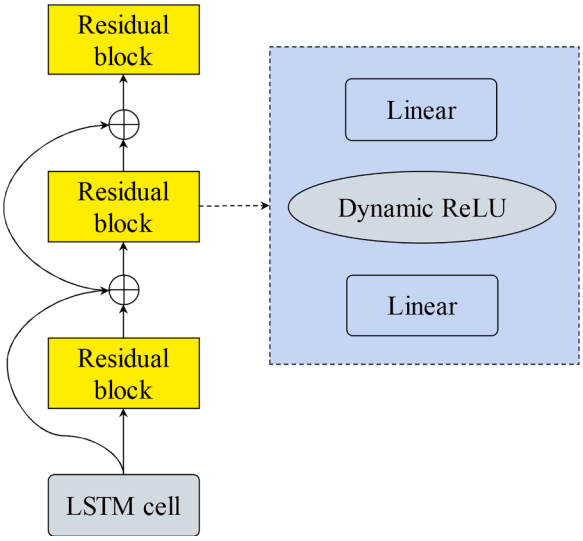

Fig. 3. Hierarchical residual connected LSTM.

$$\widetilde{h}_i^{(l)} = \text{Attention}\left(Q_i^{(l)}, K^{(l)}, V^{(l)}\right). \tag{10}$$

where $Q_i^{(l)} = W_q h_i^{(l)}$, $K^{(l)} = W_k H^{(l)}$, and $V^{(l)} = W_v H^{(l)}$ are the query, key, and value matrices, respectively, and $W_q, W_k, W_v \in \mathbb{R}^{h\times h}$ are learnable parameters.

Including extra modifications within the residual connections makes it possible for the model to reflect various designs and interrelationships in hyperscale healthcare data, hence improving the accuracy of data analysis in a complex environment.

The large model-based method can effectively capture long-term dependencies and hierarchical structures found in hyperscale healthcare data by introducing hierarchical residual connections into LSTM architecture. The vanishing gradient problem is alleviated by this type of LSTM, which also allows learning of intricate nonlinear transformations, thereby acting as a powerful building block for healthcare data fusion analysis in complex multi-sensors.

LM-HHDA incorporates temporal attention mechanisms within the LSTM layers, allowing it to capture long-term dependencies and seasonal patterns in longitudinal patient data. This enables accurate modeling of disease progression and treatment response over time.

2.3. Local feature modeling

When processing large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors, it is important to capture the details because they help us understand the patterns and relationships within this information. The multi-scale convolutional architecture in our local feature modeling module processes various types of data from multiple sensors, including ECG signals, vital signs (e.g., blood pressure, heart rate, oxygen saturation), lab results, and medical imaging data. This allows the model to capture local patterns and features across different modalities of healthcare data. The hierarchical residual connected LSTM architecture introduced in the previous subsection can capture long-term dependencies and hierarchical structures very well. However, they might not accurately represent local patterns or show fine-grained details. To enable such models to understand hyperscale healthcare data in complex multi-sensors better, we propose a convolutional neural network (CNN) -based module for local features modeling [26]. The multi-scale convolutional architecture is designed to simultaneously process and fuse local features from various sensors. This multi-sensor local feature fusion enables the model to capture intricate patterns across different data sources.

Due to their ability to detect local patterns and extract relevant features from input sources, CNNs have succeeded in different fields like computer vision [27] and natural language processing [28]. Therefore, adding another component to the large model-driven method can also consider these global aspects.

Large model-driven methods' success in healthcare data analysis depends on completeness and quality. Noise or incomplete data may distort the results of the model's forecasts. To avert these challenges, there is a need to create robust data collection and curation processes. This involves checking for errors in entries, correctly managing missing values, and ensuring that the collected data is coherent and precise.

Commonly faced with the problem of missing data, healthcare datasets considerably affect models' performance. Missing data is another source of bias and incompleteness that can affect the truthfulness and precision of model predictions. k-nearest neighbors imputation alleviates the impact of missing data by approximating the values not available from present information.

We perform a set of convolutions at multi-scales to detect local features. The convolution operation at position i with kernel size k is defined as:

$$c_i = \text{ReLU}(W_c \cdot e_{i:i+k-1} + b_c). \tag{11}$$

where $W_c \in \mathbb{R}^{k\times d\times d_c}$ and $b_c \in \mathbb{R}^{d_c}$ are learnable parameters, d_c is the number of output channels, $\cdot$ denotes the dot product, and ReLU is the rectified linear unit activation function.

Applying convolutional operation in every i^{th} position of the input sequence gives us a characteristic map $C = [c_1, c_2, ..., c_{n-k+1}]$, where each $c_i \in \mathbb{R}^{d_c}$ represents the local features captured at position i. We use multiple convolutional layers with different kernel sizes and output channels to capture local features at different scales — creating a multi-scale convolutional architecture.

However, for hyperscale healthcare data fusion analysis, input sequences can be excessively lengthy, leading to an enormous number of local features that require processing. In order to handle these high-dimensional local features efficiently and reduce computational complexity simultaneously, we propose using a pooling operation after every convolutional layer. This will downsample the feature map by selecting the most informative features within a fixed-size window.

Formally, let $C^{(l)} = \left[c_1^{(l)}, c_2^{(l)}, ..., c_{n_l}^{(l)}\right]$ be the feature map obtained after applying the l-th convolutional layer where n_l is the length of feature map. We apply max-pooling operation with window size p and stride s to get pooled feature map $P^{(l)} = \left[p_1^{(l)}, p_2^{(l)}, ..., p_{n_p}^{(l)}\right]$.

$$p_i^{(l)} = \max_{j\in[1,p]} c_{(i-1)s+j}^{(l)}. \tag{12}$$

We select the most important local features in each window using max pooling to reduce the number of features while retaining vital information.

The result of applying such an architecture is a collection $P^{(1)}, P^{(2)}, ..., P^{(L)}$ of pooled feature maps, where L is the number of convolutional layers used. In order to get these global features to work together with local ones obtained through hierarchical residual connected LSTM, we flatten every pooled feature map $P^{(l)}$ into a vector $\widetilde{p}^{(l)} \in \mathbb{R}^{n_p d_c}$ before combining them all as one large representation for our system's awareness about locality.

$$\widetilde{p} = \left[\widetilde{p}^{(1)}; \widetilde{p}^{(2)}; ...; \widetilde{p}^{(L)}\right]. \tag{13}$$

where $\widetilde{p} \in \mathbb{R}^{d_p}$ and $d_p = \sum_{l=1}^{L} n_p^{(l)} d_c^{(l)}$ is the total dimensionality of the local feature representation.

Finally, a concatenation or element-wise addition fusion mechanism can combine the global feature representation h generated from the hierarchical residual connected LSTM and the local feature representation $\widetilde{p}$.

$$f = \text{Fusion}(h, \widetilde{p}). \tag{14}$$

where $f \in \mathbb{R}^{d_f}$ is the fused feature representation, and d_f depends on the chosen fusion mechanism.

The fused characteristic representation f captures not only the worldwide hierarchical structures but also the local fine-grained patterns existing in hyperscale healthcare data, which makes the model able to carry out a more thorough and accurate analysis in complicated multi-sensors.

Our method introduces a CNN-based local feature modeling component into a large model-driven approach to capture local patterns and fine-grained details in hyperscale healthcare data. Using multi-scale convolutional architecture followed by pooling operations allows the model to learn hierarchical representations of local features efficiently while reducing computational complexity. Moreover, this algorithm fuses global and local features, which helps us gain insights into complex relationships between different parts of the dataset that were previously unknown or hard-to-see through other means – thus enabling better accuracy and more insightful healthcare data fusion analysis under such conditions. Additionally, connecting hierarchical residual LSTM architecture with this kind of foundation creates powerful frameworks suitable for conducting large-scale models through various healthcare environments involving different complexities.

2.4. Dynamic ReLU

In large model-driven hyperscale medical data analysis via complex multi-sensors, selecting an activation function is very important since this determines how well a model can capture intricate patterns and interdependencies among data. Rectified linear unit (ReLU) [29], among other traditional activation functions, has wide application in deep learning models because they are simple yet effective in dealing with vanishing gradient problems. Nevertheless, these fixed activation functions might not be the best option for processing healthcare information systems' variety and complexity, especially for large-scale models operating under complex multi-sensors.

To overcome this limitation while making our model more flexible concerning hyperscale healthcare datasets, we propose dynamic ReLU as an alternative. A dynamic ReLU adjusts its activation threshold depending on input data to represent better those fine-grained structures or relationships in healthcare domains that were not captured by previous methods.

The main concept behind dynamic ReLUs lies in learning activation thresholds according to input data rather than using fixed ones like standard ReLUs. This enables our models to adjust their behavior activations depending on different properties exhibited by inputs, thereby leading to stronger representation and wider coverage capability for capturing intricate designs.

Let $x \in \mathbb{R}^d$ be input into a dynamic ReLU activation function where we define such units as follows:

$$f(x) = \max(0, x - \alpha(x)). \tag{15}$$

where $\alpha(x) \in \mathbb{R}^d$ is the learnable activation threshold, a function of the input x. The activation threshold $\alpha(x)$ is computed using a small neural network, allowing it to adapt to the specific characteristics of the input data.

To compute the activation threshold $\alpha(x)$, we apply a global average pooling operation to the input x to obtain a summary statistic $s \in \mathbb{R}$:

$$s = \frac{1}{d}\sum_{i=1}^{d} x_i. \tag{16}$$

Next, we pass the summary static s via a fully connected network with one hidden layer to compute the activation threshold.

$$h = \text{ReLU}(W_1 s + b_1). \tag{17}$$

$$\alpha(x) = W_2 h + b_2. \tag{18}$$

where $W_1 \in \mathbb{R}^{k\times 1}$, $b_1 \in \mathbb{R}^k$, $W_2 \in \mathbb{R}^{d\times k}$, and $b_2 \in \mathbb{R}^d$ are learnable parameters, and k is the dimensionality of the hidden layer.

To support the model capture of complex patterns within hyperscale healthcare data, we further extend dynamic ReLU to work at multiple scales. Specifically, we use them on various layers' outputs in the hierarchical residual connected to LSTM and the multi-scale convolutional architecture introduced earlier.

Let $h^{(l)} \in \mathbb{R}^{d_l}$ denote the output of l-th layer in hierarchical residual connected LSTM and $p^{(l)} \in \mathbb{R}^{d_p^{(l)}}$ be the output of l-th layer in multi-scale convolutional architectures. We apply the dynamic ReLU activation function to each one of these outputs:

$$\widetilde{h}^{(l)} = \max\left(0, h^{(l)} - \alpha_h^{(l)}\left(h^{(l)}\right)\right). \tag{19}$$

$$\widetilde{p}^{(l)} = \max\left(0, p^{(l)} - \alpha_p^{(l)}\left(p^{(l)}\right)\right). \tag{20}$$

where $\alpha_h^{(l)}$ and $\alpha_p^{(l)}$ are the adaptive activation thresholds for the l-th layer of the hierarchical residual connected LSTM and the multi-scale convolutional architecture, respectively.

The dynamic ReLU activation function is further adapted to handle multi-sensor inputs, with separate activation thresholds learned for different sensor types. This sensor-specific activation allows for more nuanced processing of diverse data streams.

Furthermore, we propose a parameter-sharing mechanism across different layers to ensure that dynamic ReLU activation functions are computationally efficient and scalable for large-scale models.

$$h^{(l)} = \text{ReLU}\left(W_1 s^{(l)} + b_1\right). \tag{21}$$

$$\alpha^{(l)}\left(x^{(l)}\right) = W_2 h^{(l)} + b_2. \tag{22}$$

where $s^{(l)}$ is the summary statistic computed from the input $x^{(l)}$ at the l-th layer, and W_1, b_1, W_2, and b_2 are shared parameters across all layers.

It is more computationally efficient and suitable for complex settings of hyperscale healthcare data fusion analysis by revealing the parameters of dynamic ReLU in different layers, significantly decreasing the number of the model's learnable parameters.

With multi-scale application through hierarchical residual connected LSTM and multi-scale convolutional architecture at different levels of granularity, it is also known as a large model approach based on patterns and dependencies driven by healthcare data. In other words, our large models become capable enough to capture complex relationships within vast amounts of information in healthcare at various scales when they incorporate such an algorithm.

2.5. Relationship prediction

Predicting the correlation between medical entities with high accuracy is crucial in understanding disease progression, treatment efficacy, and patient outcomes within large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors. In our approach, the relationship prediction component attempts to discover complicated relationships inherent in healthcare data by using rich features extracted by previous components such as hierarchical residual connected LSTM, multi-scale convolutional architecture, and dynamic ReLU activation function. The relationship prediction component is enhanced to identify and predict relationships between medical entities and different sensor readings. This cross-sensor relationship analysis provides a more holistic view of the patient's health status.

After receiving fused feature representation $f \in \mathbb{R}^{d_f}$ from former modules, we propose a global attention mechanism to identify which features are most relevant for predicting relationships. Through the

attention mechanism, the model can concentrate on the key information essential for inferring relationships among medical entities while suppressing noise and irrelevant details.

In the relationship prediction component of this work, the attention mechanism is highly important in determining which features are most useful to predict a relationship between medical entities. On the other hand, one can interpret attention weights as reflecting how much a given feature contributes to prediction. This implies that if an attribute has a higher weight assigned to it, this particular attribute captured in this feature was more informative and contributed more to making an accurate prediction. Additionally, visualized attention weights such as heat maps and word clouds provide additional insights into how these models function and support communication with stakeholders regarding their use.

Let $Q \in \mathbb{R}^{d_q \times d_f}$, $K \in \mathbb{R}^{d_k \times d_f}$ and $V \in \mathbb{R}^{d_v \times d_f}$ be query matrix, key matrix and value matrices respectively where d_q, d_k and d_v are dimensionalities of query space, key space and value space. While Q, K, and V are derived from the input X through linear projections:

$$Q = XW_Q. \tag{23}$$

$$K = XW_K. \tag{24}$$

$$V = XW_V. \tag{25}$$

Where W_Q, W_K, and W_V are learnable weight matrices. These projections allow the model to transform the input into different subspaces for attention computation.

Then we compute attention scores $a \in \mathbb{R}^{d_q}$ as follows:

$$a = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right). \tag{26}$$

where softmax is the softmax function, and $\sqrt{d_k}$ is a scaling factor to prevent the attention scores from becoming large.

Each attention score a shows how significant each feature in the combined representation f is for predicting relations. To determine the resemblance between the query and every feature, we calculate the dot product of query matrix Q by key matrix K, then normalize it with the softmax function to get attention scores.

We multiply element-wise a and f to integrate attention scores into a fused feature representation.

$$\widetilde{f} = a \odot f. \tag{27}$$

where $\odot$ denotes element-wise multiplication, and $\widetilde{f} \in \mathbb{R}^{d_f}$ is the attended feature representation.

The attended feature representation $\widetilde{f}$ effectively captures the most pertinent data for predicting relationships while discarding noise and irrelevant details, enabling the model to pay attention only to the main trends and interconnections in healthcare data.

We implement a robust loss function based on the generalized cross-entropy approach to address label noise. This loss function is less sensitive to noisy labels, allowing the model to learn effectively even in label inconsistencies.

To predict relationships between medical entities, we recognize named entities in the input data using a NER module, which may be a separate pre-trained model or part of the overall architecture.

For each pair of entities (e_i, e_j), we take corresponding attended feature representations $\widetilde{f}_i$ and $\widetilde{f}_j$ from $\widetilde{f}$ given their positions in input data and then compute pairwise feature representation *gij* as concatenation of $\widetilde{f}_i, \widetilde{f}_j$ with element-wise product between them:

$$g_{ij} = \left[\widetilde{f}_i; \widetilde{f}_j; \widetilde{f}_i \odot \widetilde{f}_j\right]. \tag{28}$$

where $[\cdot;\cdot]$ denotes concatenation, and $g_{ij} \in \mathbb{R}^{3d_f}$.

The necessary information for predicting the relationship among entities e_i and e_j is contained in the pair-wise feature representation, g_{ij}. The addition of $\widetilde{f}_i \odot \widetilde{f}_j$ can model interactions and dependencies between features of different entities that are necessary for correct inference about their relationships.

We use a multi-layer perceptron (MLP) followed by a softmax layer to pass through each (e_i, e_j)'s predicted relationship label.

$$h_{ij}^{(1)} = \text{ReLU}\left(W_1 g_{ij} + b_1\right). \tag{29}$$

$$h_{ij}^{(2)} = \text{ReLU}\left(W_2 h_{ij}^{(1)} + b_2\right). \tag{30}$$

$$h_{ij}^{(L)} = \text{ReLU}\left(W_L h_{ij}^{(L-1)} + b_L\right). \tag{31}$$

$$p_{ij} = \text{softmax}\left(W_p h_{ij}^{(L)} + b_p\right). \tag{32}$$

where $W_1 \in \mathbb{R}^{d_1 \times 3d_f}$, $b_1 \in \mathbb{R}^{d_1}$, $W_2 \in \mathbb{R}^{d_2 \times d_1}$, $b_2 \in \mathbb{R}^{d_2}$, ..., $W_L \in \mathbb{R}^{d_L \times d_{L-1}}$, $b_L \in \mathbb{R}^{d_L}$, $W_p \in \mathbb{R}^{n_c \times d_L}$, and $b_p \in \mathbb{R}^{n_c}$ are learnable parameters, n_c is the number of relationship classes, and L is the number of layers in the MLP.

In order to train the component for predicting relationships, we apply a cross-entropy loss function that calculates the difference between anticipated probability distribution p_{ij} and accurate relationship label $\hat{r}_{ij}$. The classification labels in medical data context include diagnostic codes (ICD-9/10), procedure codes, and medication codes. For relationship prediction, the relationship labels represent the semantic relationships between medical entities (e.g., 'treats' between a medication and a condition, 'causes' between a condition and a symptom, or 'contraindicates' between two medications).

$$\mathscr{L} = -\sum_{(ij)\in\mathscr{P}} \sum_{k=1}^{n_c} \hat{r}_{ij}[k] \log p_{ij}[k]. \tag{33}$$

where $\mathscr{P}$ is the set of entity pairs, and $\hat{r}_{ij}[k]$ is the one-hot encoding of the true relationship label for the entity pair (e_i, e_j).

The model should be taught to minimize cross-entropy loss to predict the relationship labels between entity pairs in healthcare data. A model can learn complex non-linear decision boundaries within the feature space to detect intricate patterns and dependencies among entities through multi-layer perceptron.

We employ several regularization techniques during training to further improve the model's performance and generalization ability. These include L2 regularization on the model's parameters to prevent overfitting, dropout on the hidden layers of the MLP to improve robustness, and early stopping based on the validation set performance to avoid overtraining.

To address the critical need for interpretability in healthcare AI systems, we integrate several techniques into the LM-HHDA model to enhance its explainability. At the global level, we employ shapley additive explanation values to quantify the importance of each input feature across the entire dataset. This allows clinicians to understand which factors influence the model's predictions. We implement a local interpretability method based on local interpretable model-agnostic explanations for individual predictions. This technique approximates the complex LM-HHDA model with a simpler, interpretable model around each specific prediction, providing insights into why particular decisions were made for individual patients. Additionally, we incorporate attention visualization techniques to illustrate which parts of the input data (e.g., specific words in clinical notes or time points in sensor data) the model focuses on when making predictions. This is particularly useful for understanding how the model processes temporal and textual information.

In summary, this part of the large model-driven approach for predicting relationships relies heavily on previous components' extraction

capabilities while dealing with massive amounts of medical records from different sources. Since it considers all possible inputs together at once (pairs), a global attention mechanism can be used here, focusing only on the most relevant information necessary for making predictions about relations. Additionally, pairwise representation considers interactions between features belonging to one entity and those belonging to another entity involved in the relationship being considered. Therefore, such features are likely obtained when working with natural language processing tasks involving semantic parsing, where words often occur close together syntactically but far apart semantically within sentence boundaries. The multi-layer perceptron has softmax outputs, which means it can learn complex decision boundaries. Therefore, it can give highly accurate relationship label prediction results. Such component integration with hierarchical residual connected LSTM, multi-scale convolutional architecture dynamic ReLU activation function creates strong foundations for large size driven hyperscale healthcare data fusion analysis under complex environment settings.

3. Experiment results and analysis

3.1. Experimental setup

We perform experiments on two commonly used healthcare datasets – MIMIC-III [30] and eICU-CRD [31] – to evaluate the performance of our proposed LM-HHDA method for large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors. MIMIC-III is a large, freely available database containing de-identified health-related records for more than 40,000 patients admitted to the critical care units of Beth Israel Deaconess Medical Center between 2001 and 2012. The eICU-CRD is a multi-center database with de-identified health-related data for over 200,000 ICU admissions across the United States in 2014–2015. The training and inference of the LM-HHDA model were performed on a computing cluster with 8 NVIDIA Tesla V100 GPUs, each with 32 GB of memory. The training process was distributed across the GPUs using data parallelism techniques. The average training time for the model on the MIMIC-III dataset was approximately 48 h, while on the eICU-CRD dataset, it took around 72 h.

Our preprocessing pipeline includes (1) removal of duplicate records, (2) standardization of medical terminologies using SNOMED CT, (3) normalization of numerical features using z-score scaling, and (4) tokenization of clinical notes using a domain-specific tokenizer trained on medical corpora.

The use of healthcare datasets entails the consideration of biases and fairness issues. Like many other healthcare datasets, MIMIC-III and eICU-CRD may have inherent biases arising from various factors such as patient characteristics, data collection methods, and institutional practices. Sometimes, these partialities can cause variation in the model's performance across subgroups. It is necessary to perform extensive data analysis to recognize and interpret potential areas of bias to minimize them, thereby assuring fairness. Bias removal through data balancing improves model fairness. Integrating fairness metrics and constraints in the training process would also facilitate fair predictions, thus enhancing equity in healthcare applications.

Our preprocessing involves tokenizing clinical notes, normalizing medical entities, and extracting relevant features from both datasets. We divide these sets into three parts at a ratio of 70:10:20, representing the training, validation, and testing sets, respectively. The training set is then used to train the proposed LM-HHDA method, while hyperparameters are tuned using the validation set before evaluating the final performance with the test set.

The preprocessing pipeline for the MIMIC-III and eICU-CRD datasets involved several steps. First, we performed data cleaning by removing incomplete or inconsistent records. Next, we applied text normalization, e.g., lowercasing, tokenization, and removal of stop words and special characters, to standardize the unstructured clinical notes. Additionally, we used NER models to identify and extract relevant medical entities, such as diseases, medications, and procedures. The extracted entities were mapped to standardized medical codes using domain-specific ontologies. Finally, we performed feature scaling and normalization on the structured data to ensure comparability across different features.

Table 1 presents the key parameters used in our experiments for the LM-HHDA model. These parameters were determined through extensive experimentation on the validation set to optimize model performance while maintaining computational efficiency.

We compare the LM-HHDA method against six state-of-the-art baselines: GNN-MNER [32], MedSINE [33], MMBERT [34], GPDMiner [35], Entity-BERT [36], and GBPE [37]. These methods present different ways to approach healthcare data fusion analysis, such as graph neural networks, supervised learning, transformer-based models, and entity recognition frameworks.

Precision, recall, and F1 score are our experiments' evaluation metrics. For multi-classification tasks, we report macro-averaged along with micro-averaged F1 scores. Macro-averaged entails taking the average F1 score for each class. Simultaneously, micro-averaged computes it by counting true positives (TP), false negatives (FN), and false positives (FP) across all classes.

3.2. Results analysis

To demonstrate the robustness of the proposed method, we conducted ten independent runs for each method and reported the mean and standard deviation of the F1 scores in Table 2.

Table 2 shows the average F1 scores and their standard deviations of LM-HHDA and six other methods on MIMIC-III and eICU-CRD datasets. These results are based on ten independent runs for each method to ensure statistical robustness. LM-HHDA consistently outperforms all other methods on both datasets, demonstrating the highest F1 scores across all metrics with the lowest standard deviations.

GBPE is the best-performing baseline on the MIMIC-III dataset, achieving a macro-averaged F1 score of 0.868 ± 0.003 and a micro-averaged F1 score of 0.884 ± 0.002. However, LM-HHDA surpasses these scores by 2.3 % and 1.9 %, respectively, with scores of 0.891 ± 0.002 for macro-averaged F1 and 0.903 ± 0.001 for micro-averaged F1. The lower standard deviations of LM-HHDA (± 0.002 and ± 0.001) compared to GBPE (± 0.003 and ± 0.002) indicate that our method not only achieves higher performance but also demonstrates greater consistency across multiple runs.

Similarly, for the eICU-CRD dataset, LM-HHDA outperforms GBPE with improvements of 2.6 % for the macro-averaged F1 score (0.876 ± 0.003 vs 0.850 ± 0.004) and 2.2 % for the micro-averaged F1 score (0.892 ± 0.002 vs 0.870 ± 0.003). Again, the lower standard deviations of LM-HHDA suggest more stable performance across different runs.

These results demonstrate that LM-HHDA achieves superior performance in terms of F1 scores and exhibits greater consistency and robustness compared to existing methods. The combination of higher mean scores and lower standard deviations highlights the effectiveness of our approach in handling the complexities of hyperscale healthcare

Table 1
Key parameters for LM-HHDA model.

Component	Parameter	Value
Hierarchical residual connected LSTM	Number of LSTM layers	3
	Hidden state size	256
	Dropout rate	0.1
Dynamic ReLU	Hidden layer size	64
	Activation function	tanh
Attention mechanism	Number of attention heads	8
	Key/Query/Value dimension	64
Training	Batch size	32
	Learning rate	0.001
	Number of epochs	100
	Early stopping patience	10
Model size	Total trainable parameters	5.2M

Table 2
Macro-averaged and micro-averaged F1 scores (Mean ± Standard Deviation) of LM-HHDA and The Baseline Methods on The MIMIC-III and eICU-CRD Datasets.

Method	MIMIC-III dataset		eICU-CRD dataset	
	Macro F1	Micro F1	Macro F1	Micro F1
GNN-MNER	0.793 ± 0.008	0.811 ± 0.007	0.776 ± 0.009	0.794 ± 0.008
MedSINE	0.814 ± 0.007	0.829 ± 0.006	0.798 ± 0.008	0.815 ± 0.007
MMBERT	0.832 ± 0.006	0.847 ± 0.005	0.815 ± 0.007	0.833 ± 0.006
GPDMiner	0.846 ± 0.005	0.860 ± 0.004	0.829 ± 0.006	0.847 ± 0.005
Entity-BERT	0.857 ± 0.004	0.871 ± 0.003	0.840 ± 0.005	0.858 ± 0.004
GBPE	0.868 ± 0.003	0.884 ± 0.002	0.850 ± 0.004	0.870 ± 0.003
LM-HHDA	0.891 ± 0.002	0.903 ± 0.001	0.876 ± 0.003	0.892 ± 0.002

data in multi-sensor environments.

The results demonstrate that LM-HHDA better understands perplexing structures and associations in hyperscale medical data. LM-HHDA can learn from various unstructured health data because it combines global attention mechanisms with dynamic rectified linear unit activation functions, multi-scale convolutional architectures, hierarchical residual connected LSTMs, and large models. As a result, the method creates more accurate representations, thus enhancing its capacity to improve performance in healthcare data fusion analysis tasks.

For further examination of the LM-HHDA's ability to predict specific healthcare-related categories, we have performed experiments on multi-class classification tasks with the MIMIC-III and eICU-CRD datasets. Primary diagnosis, medication, and procedure categories are predicted in these tasks from a patient's clinical notes and structured data.

Fig. 4(a) shows how well LM-HHDA performs in classifying the MIMIC-III dataset against other methods. Among all three classification tasks, LM-HHDA has the highest F1 scores, thus considerably outperforming baselines. For instance, in the primary diagnosis prediction task, GBPE, the best baseline, achieves an F1 score of only 0.886 while our model reaches 0.912 (2.6 % higher than GBPE). Similarly, for medication and procedure prediction tasks, GBPE has F1 scores 0.918 and 0.899, respectively, but LM-HHDA achieves them with values equal to 0.897 (2.4 % better than GBPE) and 0.883 (2.2 % more accurate than GBPE) correspondingly. The eICU-CRD dataset's classification performance is depicted in Fig. 4(b). Across all classification tasks, LM-HHDA consistently outperforms the baselines, which means it is robust and can be applied to various healthcare datasets. Regarding predicting primary diagnoses, LM-HHDA outperforms GBPE by 2.8 % as its F1 score amounts to 0.897. In the case of the medication prediction task, this algorithm achieves an F1 score equal to 0.881 (GBPE's result was lower by 2.5 %), while for the procedure prediction task – 0.867 (GBPE had only 0.2 % higher value).

These results show how important it is to use LM-HHDA to forecast particular health categories that are essential in decision support systems and personalized medicine tools. What differentiates LM-HHDA from other algorithms is its capability to extract deep features from complicated and diverse healthcare records, thus recognizing subtle interdependencies among various classes, which indicates better classification accuracy.

In order to study the LM-HHDA model's every single component contribution, we did ablation experiments on eICU-CRD and MIMIC-III datasets. We removed one part each time and measured its effect on LM-HHDA performance together with different versions of it that were obtained by getting rid of some other parts, including hierarchical residual connected LSTM (HR-LSTM), dynamic ReLU activation function (D-ReLU), global attention mechanism(GA) and multi-scale convolutional architecture(MS-CNN).

For the MIMIC-III dataset, Fig. 5(a) displays results from the ablation study. Performance decrease was observed after eliminating any component, meaning all modules are important in this model. Hierarchical residual connected LSTM and multi-scale convolutional architecture have the greatest impact on performance, showing maximum reduction in macro F1 score averages by 2.1 %and 1.8 %, respectively, while dynamic ReLU activation function as well as global attention mechanism also made significant contributions since they resulted in a decrease by 1.4 % and 1.1 % in macro-averaged F1 score respectively.

The ablation study results from the eICU-CRD dataset can be found in Fig. 5(b). When compared with MIMIC-III datasets where the same experiment was done too but only this time round using a different data source, we notice that there is still similarity between them such that removal of each part leads to a drop in performance, thus confirming the proposed modules' effectiveness within LM-HHDA, model for both cases. However, when it comes to what impacts most on these two setups, among others like macroscale CNNs or even d-ReLUs, then again, we find out through average F-scores, which were reduced by 2.3 %for HR-LSTMs against MS-CNNs having decreased them by 1 0.9 %.

To gain insights into the performance of LM-HHDA and identify potential areas for improvement, we conduct an error analysis by examining the confusion matrices on the MIMIC-III and eICU-CRD datasets. Fig. 6 presents the confusion matrix for the primary diagnosis prediction task on the MIMIC-III dataset. The confusion matrix reveals that LM-HHDA achieves high accuracy for most classes, with few exceptions. The model tends to misclassify certain classes, such as "Respiratory system" and "Digestive system," which share similar symptoms and clinical presentations.

In order to understand the performance of LM-HHDA and areas that can be improved, we carried out an error analysis by looking at confusion matrices on MIMIC-III and eICU-CRD datasets. The primary diagnosis prediction task's confusion matrix for the MIMIC-III dataset is

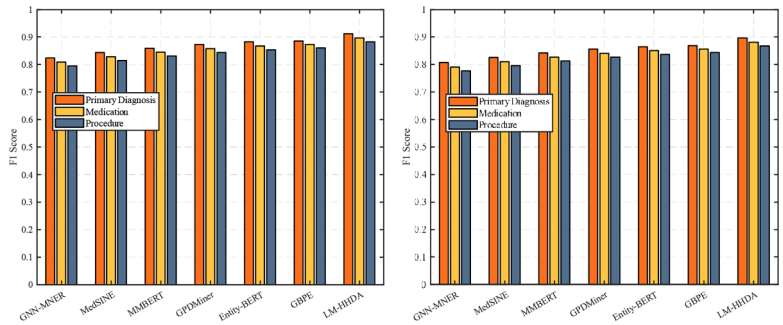

Fig. 4. Classification performance of LM-HHDA and the baseline methods on the MIMIC-III and eICU-CRD datasets.

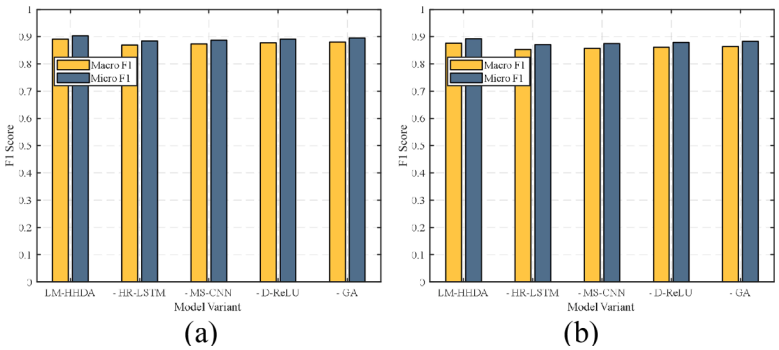

Fig. 5. Ablation study results on the MIMIC-III and eICU-CRD datasets.

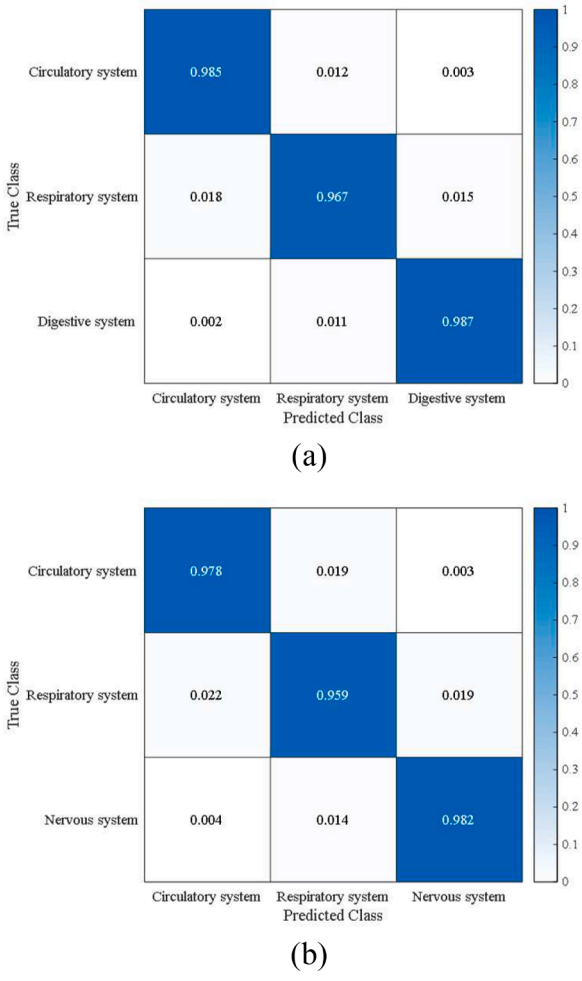

Fig. 6. Confusion matrices on the MIMIC-III and eICU-CRD datasets.

shown in Fig. 6(a). It can be seen from the confusion matrix that most classes are highly accurate except for a few. The model frequently misclassifies some of them, for example, diseases with similar symptoms and clinical presentations like "Respiratory system" and "Digestive system."

Fig. 6(b) displays the confusion matrix used in this task on the eICU-CRD dataset, where it also shows high accuracy rates among many different categories while still making some mistakes between closely related ones, such as between "Circulatory system" and "Respiratory system" which were classified wrongly by LM-HHDA just like it did with MIMIC-III dataset.

These error analyses reveal how difficult it is to differentiate between closely related classes during healthcare data fusion analysis tasks. The overlaps in symptoms and clinical presentations among various diseases may lead to wrong classifications made by models, including LM-HHDA and other complexities involved within healthcare data being heterogeneous.

We introduce noise into unstructured clinical notes to simulate heterogeneity by replacing some structured data proportions with random missing values. The extent to which heterogeneity is simulated depends on the percentage missingness and the amount of noisy signal added to records.

On the MIMIC-III dataset, we illustrate how much difference there can be between various methods used in dealing with healthcare information when faced with increasing levels Fig 7(a).

To analyze hyperscale healthcare data, the process has to be scalable enough. Scalability can be evaluated by modifying LM-HHDA. They should then judge its scalability using training time and model performance.

Fig. 8(a) compares the scalability analysis between LM-HHDA and baseline methods on the MIMIC-III dataset. In all cases, as expected, the time taken for training increases with the size of training data used for all algorithms, but not equally well. Fig. 8(b) shows that when compared against baseline methods, LM-HHDA performs better because its training time does not grow rapidly when more input samples are employed while keeping other conditions constant, such as the type or nature of an algorithm applied to medical records stored within any given hospital database among others which may affect this metric too.

To better assess the clinical relevance of the LM-HHDA model, we have expanded our evaluation to include healthcare-specific metrics. In addition to the previously reported accuracy, precision, recall, and F1 scores, we incorporate the following metrics:

- Area under the precision-recall curve (AUPRC): This metric is particularly useful for imbalanced datasets, common in healthcare where certain conditions may be rare.
- Net reclassification improvement (NRI): This metric assesses the improvement in risk stratification compared to baseline methods.
- Calibration curves and expected calibration error (ECE): These evaluate how well the model's predicted probabilities align with observed frequencies, which is crucial for risk assessment in clinical settings.
- Clinical decision curve analysis (DCA): This metric evaluates the clinical usefulness of the model across different decision thresholds.

Table 3 presents the results of these additional metrics for LM-HHDA and baseline methods on both MIMIC-III and eICU-CRD datasets.

LM-HHDA achieves the highest AUPRC (0.894), indicating superior performance in identifying rare but clinically significant events. This is particularly important in healthcare applications where false negatives can have severe consequences. The positive NRI value (0.287) for LM-HHDA suggests a significant improvement in risk stratification

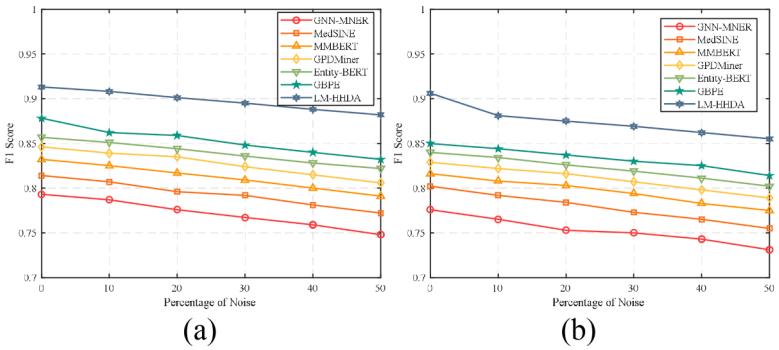

Fig. 7. Impact of data heterogeneity on the performance of LM-HHDA and the baseline methods on the MIMIC-III and eICU-CRD datasets.

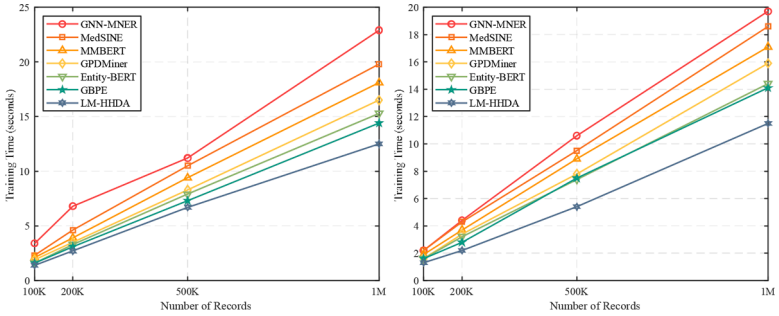

Fig. 8. Scalability analysis of LM-HHDA and the baseline methods on the MIMIC-III and eICU-CRD datasets.

Table 3
Healthcare-specific metrics for LM-HHDA and baseline methods.

Method	AUPRC	NRI	ECE	DCA (Net benefit at 30 % threshold)
GNN-MNER	0.763	–	0.089	0.142
MedSINE	0.789	0.085	0.076	0.165
MMBERT	0.815	0.132	0.062	0.183
GPDMiner	0.837	0.178	0.055	0.201
EntityBERT	0.852	0.203	0.048	0.219
GBPE	0.871	0.235	0.041	0.237
LM-HHDA	0.894	0.287	0.033	0.256

compared to the best-performing baseline (GBPE). This indicates that LM-HHDA more accurately classifies patients into appropriate risk categories, potentially leading to more targeted interventions. LM-HHDA shows the lowest ECE (0.033), indicating that its predicted probabilities closely align with observed frequencies. This is crucial for reliable risk assessment in clinical decision-making. At a typical decision threshold of 30 %, LM-HHDA provides the highest net benefit (0.256), suggesting that it offers the most clinical utility among the compared methods. This implies that using LM-HHDA for decision support could lead to better patient outcomes than treating all patients or using other predictive models.

While we have not directly compared our model to general large language models like ChatGPT, we can highlight scenarios where our LM-HHDA method is likely to outperform:

- Multi-modal data integration: Our model can process and integrate data from various sensors and modalities, while ChatGPT is primarily designed for text input.
- Domain-specific knowledge: LM-HHDA is fine-tuned on healthcare datasets, allowing it to understand complex medical terminology and relationships that general models might struggle with.
- Temporal dependencies: Our model's LSTM components are specifically designed to capture long-term dependencies in time-series healthcare data, which is crucial for predicting patient outcomes.
- Structured output: LM-HHDA can produce structured predictions (e.g., specific diagnosis codes), while ChatGPT generates free-form text that may not adhere to healthcare standards.

These advantages make LM-HHDA more suitable for specific healthcare tasks that require precise, multi-modal analysis and structured outputs.

The experiments show how well LM-HHDA performs across different healthcare data fusion analysis tasks such as classification, prediction, and addressing data heterogeneity. Its scalability, interpretability, and robustness make it an attractive option for dealing with hyperscale healthcare data analytics in complicated multi-sensors. The fact that LM-HHDA can learn deep expressive representations from various unstructured data sources through large models, multi-scale convolutional architectures, hierarchical residual connected LSTMs, dynamic ReLU activation functions along with global attention mechanisms has resulted in enhanced decision support capabilities besides the best performance so far achieved.

4. Conclusion

This paper proposed a large model-driven method for processing hyperscale healthcare data fusion analysis in complicated scenarios named LM-HHDA. Our strategy involved utilizing powerful large models, hierarchical residual connected LSTMs, multi-level convolutional structures, dynamic ReLU activation functions, and global attention mechanisms to effectively represent intricate designs or concepts among hyperscale medical information. The experimental results prove that the proposed LM-HHDA method outperforms baseline methods across different tasks of healthcare data fusion analysis, e.g., classification, prediction, and dealing with heterogeneity in data types. Regarding accuracy, precision, recall, and F1 score, it was found that LM-HHDA consistently achieved higher scores than baseline methods, thus demonstrating its capability to learn from various sources richly and expressively.

The proposed LM-HHDA method performs better in fusing and

analyzing data from complex multi-sensor healthcare environments. Future work could focus on extending the method to handle real-time sensor data streams and improving the fusion of multimodal sensor inputs.

CRediT authorship contribution statement

Jianhui Lv: Writing – original draft, Methodology, Conceptualization. **Byung-Gyu Kim:** Investigation, Data curation. **B.D. Parameshachari:** Software, Resources, Methodology, Investigation. **Adam Slowik:** Writing – review & editing, Writing – original draft, Software, Data curation. **Keqin Li:** Writing – review & editing, Resources, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Granted No 62202247.

Data availability

Data will be made available on request.

References

[1] K. Batko, A. Slezak, The use of big data analytics in healthcare, J. Big Data 9 (1) (2022) 3.
[2] W.S. Gan, K.X. Hu, G.S. Huang, W.C. Chien, H.C. Chao, W.Z. Meng, Data analytic for healthcare cyber physical system, IEEE Trans. Netw. Sci. Eng. 10 (5) (2023) 2490–2502.
[3] X.H. Lyu, S. Rani, S. Manimurugan, C. Maple, Y.H. Feng, A deep neuro-fuzzy method for ECG big data analysis via exploring multimodal feature fusion, IEEE Trans. Fuzzy Syst (2024), https://doi.org/10.1109/TFUZZ.2024.3416217 early access.
[4] H. Elayan, M. Aloqaily, M. Guizani, Sustainability of healthcare data analysis IoT-based systems using deep federated learning, IEEe Internet. Things. J. 9 (10) (2022) 7338–7346.
[5] A. Raza, K.P. Tran, L. Koehl, S.J. Li, AnoFed: adaptive anomaly detection for digital health using transformer-based federated learning and support vector data description, Eng. Appl. Artif. Intell. 121 (2023) 106051.
[6] L. Zhang, K.J. Zhang, H.W. Pan, SUNet plus plus: a deep network with channel attention for small-scale object segmentation on 3D Medical Images, Tsinghua Sci. Technol. 28 (4) (2023) 628–638.
[7] X.L. Chen, H.R. Xie, X.H. Tao, F.L. Wang, M.M. Leng, B.Y. Lei, Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics, Artif. Intell. Rev. 57 (4) (2024) 91.
[8] Z.W. Ning, J.Y. Wu, Y.D. Ding, Y. Wang, Q.K. Peng, L.Y. Fu, BertNDA: a model based on graph-bert and multi-scale information fusion for ncRNA-disease association prediction, IEEE J. Biomed. Health Inform. 27 (11) (2023) 5655–5664.
[9] H. Ong, J. Ong, R.B.K. Cheng, C.L. Wang, M.R. Lin, D. Ong, GPT technology to help address longstanding barriers to care in free medical clinics, Ann. Biomed. Eng. 51 (9) (2023), 1906–1909.
[10] S. Bengesi, H. El-Sayed, M.K. Sarker, Y. Houkpati, J. Irungu, T. Oladunni, Advancements in Generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers, IEEe Access. 12 (2024) 69812–69837.
[11] Y.F. Chen, Z. Huang, M.H. Hu, D.S. Li, C.J. Wang, F. Liu, X.C. Lu, Decoupled two-phase framework for class-incremental few-shot named entity recognition, Tsinghua Sci. Technol. 28 (5) (2023) 976–987.
[12] C. Peng, X. Yang, Z.H. Yu, J. Bian, W.R. Hogan, Y.H. Wu, Clinical concept and relation extraction using prompt-based machine reading comprehension, J. Am. Med. Inform. Assoc. 30 (9) (2023) 1486–1493.
[13] A. Zafar, S.K. Sahoo, H. Bhardawaj, A. Das, A. Ekbal, KI-MAG: a knowledge-infused abstractive question answering system in medical domain, Neurocomput 571 (2024) 127141.
[14] Q. Guo, S. Cao, Z. Yi, A medical question answering system using large language models and knowledge graphs, Int. J. Intell. Syst. 37 (11) (2022) 8548–8564.
[15] T. Yang, Y.F. Deng, B. Yu, Y.H. Qian, J.H. Dai, Local feature selection for large-scale data sets with limited labels, IEEE Trans. Knowl. Data Eng. 35 (7) (2023) 7152–7163.
[16] G.X. Duan, J.J. Chen, Y.Y. Zhou, X.Y. Zheng, Y.X. Zhu, Large Language Model Inference Acceleration Based on Hybrid Model Branch Prediction, Electronics 13 (7) (2024) 1376. Apr.
[17] M. Saul, S. Rostami, Assessing performance of artificial neural networks and re-sampling techniques for healthcare datasets, Health Inform. J. 28 (1) (2022), 14604582221087109.
[18] T. Ishikawa, T. Yakoh, H. Urushihara, An NLP-Inspired Data Augmentation Method for Adverse Event Prediction Using an Imbalanced Healthcare Dataset, IEEe Access. 10 (2022) 81166–81176.
[19] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, J. Qadir, Explainable, trustworthy, and ethical machine learning for healthcare: a survey, Comput. Biol. Med. 149 (2022) 106043.
[20] Q. Zhao, H.R. Xu, J.Q. Li, F.A. Rajput, L.Y. Qiao, The Application of Artificial Intelligence in Alzheimer's Research, Tsinghua Sci. Technol. 29 (1) (2024) 13–33.
[21] G. Veena, V. Kanjirangat, D. Gupta, AGRONER: an unsupervised agriculture named entity recognition using weighted distributional semantic model, Expert Syst. Appl. 229 (2023) 120440.
[22] M.B. Shishehgarkhaneh, R.C. Moehler, Y.H. Fang, A.A. Hijazi, H. Aboutorab, Transformer-based named entity recognition in construction supply chain risk management in Australia, IEEe Access. (2024) 41829–41851.
[23] J.X. Zhang, L.X. Zhang, Y.X. Sun, W. Li, R.G. Quhe, Named entity recognition in the perovskite field based on convolutional neural networks and MatBERT, Comput. Mater. Sci. 240 (2024) 113014.
[24] C. Qin, L.M. Chen, Z.T. Cai, M. Liu, L. Jin, Long short-term memory with activation on gradient, Neural Netw. 164 (2023) 135–145.
[25] C. Fang, Y.K. Gong, X.B. Ming, L.M. Qin, Tool wear prediction based on LSTM and deep residual network, Int. J. Pattern Recognit. Artif. Intell. 38 (5) (2024).
[26] B.P. Xiong, W.S. Chen, Y.X. Niu, Z.H. Gan, G.J. Mao, Y. Xu, A Global and Local Feature fused CNN architecture for the sEMG-based hand gesture recognition, Comput. Biol. Med. 166 (2023) 107497.
[27] Y.K. Liu, R. Cao, S.P. Xu, L. Deng, A deep learning-based method for structural modal analysis using computer vision, Eng. Struct. 301 (2024) 117285.
[28] T.F. Lee, Y.W. Hsieh, P.Y. Yang, C.H. Tseng, S.H. Lee, J.C. Yang, L.Y. Chang, J. M. Wu, C.D. Tseng, P.J. Chao, Using meta-analysis and CNN-NLP to review and classify the medical literature for normal tissue complication probability in head and neck cancer, Radiat. Oncol. 19 (1) (2024) 5.
[29] Y.S. Xu, H.Z. Zhang, Convergence of deep ReLU networks, Neural Netw. 571 (2024) 127174.
[30] W. Yang, H. Zou, M. Wang, Q. Zhang, S.D. Li, H.Y. Liang, Mortality prediction among ICU inpatients based on MIMIC-III database results from the conditional medical generative adversarial network, Heliyon. 9 (2) (2023) e13200.
[31] Y.F. She, L.M. Zhou, Y.D. Li, Interpretable machine learning models for predicting 90-day death in patients in the intensive care unit with epilepsy, Seizure-Eur. J. Erilep. 114 (2024) 23–32.
[32] Y.C. Gong, X.Q. Lv, Z. Yuan, X.D. You, F. Hu, Y.Z. Chen, GNN-based multimodal named entity recognition, *Comput. J.*, early access (2024), https://doi.org/10.1093/comjnl/bxae030.
[33] M.Y. Landolsi, L.B. Romdhane, L. Hlaoua, Hybrid medical named entity recognition using document structure and surrounding context, J. Supercomput. 80 (4) (2024) 5011–5041.
[34] L. Fu, Z.Q. Weng, J.H. Zhang, H.H. Xie, Y.Q. Cao, MMBERT: a unified framework for biomedical named entity recognition, Med. Biol. Eng. Comput. 62 (1) (2024) 327–341.
[35] Y.J. Park, G.J. Yang, C.B. Sohn, S.J. Park, GPDminer: a tool for extracting named entities and analyzing relations in biological literature, BMC Bioinform 25 (1) (2024) 101.
[36] W.J. Lu, J.H. Jiang, Y.X. Shi, X.W. Zhong, J. Gu, L.X. Huangfu, M. Gong, Application of Entity-BERT model based on neuroscience and brain-like cognition in electronic medical record entity recognition, Front. Neurosci. 17 (2023) 1259652.
[37] H. Ming, J.Y. Yang, F. Gui, L.L. Jiang, N. An, Few-shot nested named entity recognition, Knowl.-Based Syst 293 (2024) 111688.