International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I8P103
Volume 74 Issue 8, 37-49, August 2026

Original Article

IOMT Source- Fuse: Cross-Attention Driven Latent Multi-Source Integration Framework

R. Venkatesan[1], A. Saravanan[2]

[1,2]*Department of Computer Science and Engineering, SRIHER, Chennai, India.*

[1]*Corresponding Author : u022401153@sriramachandra.edu.in*

Received: 05 January 2026 Revised: 10 June 2026 Accepted: 24 June 2026 Published: 29 August 2026

Abstract - *IoMT systems are in high demand in the global market as the healthcare domain is a critically important factor in national development and citizens' well-being. Growing need requires optimisation in various stages of IoMT systems, which makes the system scalable and economical in practical deployment. One of the primary concerns in the IoMT systems is the data handling capacity of the overall system. In the IoMT system, prediction of the patient's condition is the primary criterion that requires high accuracy and precision. The accuracy and precision highly rely on the amount of data that the analysis provides, for efficient results. Therefore, it is a prime notion to consider the vast amount of patient data with dimensionality reduction without the loss of information. Data fusion is an efficient model that ensures data integrity by providing dimensionality reduction of information without compromising the original data. The proposed work utilises the data fusion approach for the IoMT system by fusing the input data at the feature level using the Variational Autoencoder and fusing the latent representation by the Multi-Head Cross-Attention model, which overcomes the limitations of conventional concatenation. The proposed work is validated using the EHR, Imaging, and IoT sensor data from Kaggle, achieving 94.6% accuracy with a high AUC score of 97.3%. The proposed work compared the prediction results with a single modality, where the data fusion predicted 2.2% higher compared to other single modality predictions.*

Keywords - *IoMT, Data Fusion, Variational Autoencoder, Good Health and Well-Being (SDG 3), Multi-Head Attention.*

1. Introduction

The tremendous growth in the health care systems in fields like medical devices and equipment, pharmaceuticals, medical services, and healthcare facilities is meant to provide productive and accurate diagnosis for the patients' welfare. The increase in the population and technologies requires advancements in remote monitoring, real-time data analysis and improved patient outcomes. In this way, the Internet of Things (IoT) is adopted in the healthcare system, enabling proactive decision-making based on the patient's health condition that satisfies the Good Health and Well-Being of the SDG goals (SDG-3). The intervention of IoT makes the healthcare system more effective by allowing real-time monitoring of patients and communicating the information to the hospital.

There is a significant growth experienced in the IoMT market, where the projection size is expected to be between \$97.73 billion and \$286.77 billion by the end of 2025. In the future, the market is expected to grow to \$2.2 trillion by 2034, driven by the Compound Annual Growth Rate (CAGR). It is seen that more than 75% of the hospital management are interested in the adoption of IoMT systems, which makes their healthcare services and economic status scalable. The development in IoMT starts with the interest in wearable devices that track various health benefits for self-care. IoMT systems incorporate various medical sensor devices, equipment, real-time monitoring systems, implant systems, medical wearable devices, cameras, etc., to obtain the patient's condition regularly.

The expansion in the IoMT market are due to the factors like improvement in the communication technologies that produces the seamless connectivity between the patient and hospital using the 5G systems, telemedicine and virtual care helps in availing the Doctors consultation over Internet without the actual in presence, growing rate of chronic diseases among the population, health monitoring is made ease using the smart wearable devices and sensors that senses and actuates the health monitoring systems, and adoption of the history of the patient by maintaining the Electronic Health Records (EHR).

Effective healthcare systems require the efficient collection of patient information on a regular and as-needed basis. Based on the disease the patient is affected by, there are multiple data points needed for the health professional to make an accurate prediction through the online consultation process.

The data required are of different forms, such as physiological systems, environmental data, clinical data, imaging data, etc.

Thus, the IoMT system needs to address the handling efficiency in terms of enormous information and heterogeneous data collection from the patient side, which is the primary goal to make the overall health monitoring system operate efficiently. Instead of considering all the information that reduces the efficiency of the prediction analysis, the important features need to be captured, and global attention must be paid to them for the diagnosis of the patient's condition.

Based on this objective, the work proposes data fusion as the precise solution to overcome the data handling issues in the IoMT system that requires the efficient representation of patient data in a latent space dimension. The variational autoencoder-based cross-attention model is proposed for data fusion in IoMT systems, enabling feature-level fusion of patient data from heterogeneous inputs and making predictions more accurate and reliable without compromising overall system performance.

2. Related Content

The Internet of Medical Things (IoMT) has evolved into an innovative prototype in current healthcare, flawlessly linking wearable devices, medical sensors, intelligent health information systems, and edge-cloud computing frameworks to suggest constant patient monitoring, investigative support, and data-driven medical decision-making. It addresses crucial healthcare requirements such as early disease detection, long-term chronic illness supervision, and personalised treatment delivery. A key enabler of IoMT's potential is the capacity to incorporate diverse data sources, including physiological measurements, medical imaging, inertial/motion data, environmental and contextual sensors, and patient behavioural patterns, into trustworthy and clinically related insights.

Data fusion lies at the hub of this integration, combining various inputs to generate precise, rational, and actionable information. Modern advances in artificial intelligence, mainly deep learning, self-attention, and cross-attention mechanisms, have drastically prolonged IoMT capabilities, improving predictive analytics, variance detection, and multimodal health monitoring while maintaining strict privacy and safety standards. This literature survey reviews key research published between 2020 and 2024, highlighting the novelty of multimodal fusion architectures, attention-based incorporation methods, and privacy-enhanced data processing. It also examines varied healthcare applications, including posture estimation, Activity Of Daily Living (ADL) analysis, sleep quality assessment, and pandemic-related data fusion, though identifying open challenges such as computational efficiency, data confidentiality, and robust performance in real-world medical environments.

2.1. Healthcare Applications using IoMT-Enabled Data Fusion

The author in [6] spot challenges such as interoperability, latency, and cyber vulnerabilities, introduces a nomenclature of security threats and latent counteract actions, affords a complete outline of IoMT, and stresses data fusion as a centre enabler for strong healthcare decision-making. In the same way, [4] underline the addition of AI and IoMT's technological advances for predictive healthcare. To progress the precision of musculoskeletal health assessments for posture prediction, utilising multi-sensor data [7], an IoMT-enabled fusion model was proposed. Building on [9], expand this for assisted living environments, facilitating practical interference for elderly care by utilising multimodal fusion of Activities of Daily Living (ADL) data.

To address the limitations of battery life and processing ability [8], focus on lightweight, computationally efficient fusion approaches and highlight brilliant sensor data selection and dimensionality reduction to optimise performance tailored for wearable IoMT devices. Insomnia identification and treatment variation by possible appliance enhanced sleep stage classification accuracy for sleep monitoring, integrating physiological signals and environmental parameters are discussed in [10] designed an IoT-enabled data fusion framework is proposed in [14], which proposes privacy-enhanced algorithms that integrate patient symptom data, testing results, and geospatial tracking to help community health establishments as maintaining confidentiality the COVID-19-specific data fusion challenges in IoMT

2.2. Methodological Advances in IoMT Data Fusion

For smart clinical data diffusion, initiating a multi-mode blend model that supports high-speed, consistent communication in telemedicine, as surveyed in [11], the integration of Industrial IoT and IoMT is discussed. In [12], a hybrid IoMT-based innovative healthcare monitoring framework is recommended that merges adaptive wavelet entropy for feature extraction with an improved Recurrent Neural Network (RNN) for chronological modelling. The wavelet-based preprocessing stage effectively demises and captures salient features from non-stationary biomedical signals such as ECG and PPG, enabling healthy handling of physiological variability.

Adaptive wavelet entropy assists compact and discriminative attribute representation, which is consequently developed through deep feature fusion. The enhanced RNN then models temporal dependencies in the processed signals, thereby improving the detection accuracy for irregular physiological events in real-time healthcare monitoring circumstances. In [18], the authors compared the advanced AI-based data fusion techniques and classical statistical methods and offered a more exhaustive review. Whereas [17] commented on the above reference via conferencing data handling pipelines, stressing preprocessing, normalisation, and feature alignment for fusion effectiveness. [13] Discussed the usage of IoMT data fusion techniques beyond hospital or home settings, like maritime healthcare contexts.

2.3. Attention Mechanisms and Cross-Modal Fusion for IoMT

In the process of IoMT data integration, deep multimodal fusion methods have emerged as a key enabler for dealing with heterogeneous healthcare data streams. The author in [15] reviews such systems, recognising complex, transformer-based, and graph-based architectures as principal frameworks for running various sensor and imaging inputs. Extending on this, work [16] authors initiate a multi-sensor fusion approach based on self-attention mechanisms, enabling dynamic adjustment of the weight of each sensor's input based on environmental consequences-an indispensable potential for adaptive and personalised healthcare monitoring.

Modern development in cross-attention architectures, initially started for computer visualisation, has revealed emerging utilisation in the IoMT framework. Transformer architectures and cross-attention blocks (CrossViT, CAT, CrossFuse, CCNet) are progressively more tailored to multimodal biomedical tasks. The facility to model long-range dependencies and flexible modality connections makes transformers smart for extended monitoring sequences and heterogeneously sampled modalities. In [19, 20], suggest criss-cross and cross-scale attention models, striking spatial attribute interface and depiction. Likewise, authors in [21-23] acclimatise cross-attention for multi-modal medical data integration, reaching enhanced configuration among visual and non-visual modalities. In the work [24], the author described CrossFuse, a cross-attention-based infrared-visible image fusion framework, with direct potential for adaptation to medical imaging fusion tasks such as MRI-PET integration.

2.4. Recent Development of Data Fusion in IoMT Systems

The data fusion produces data integrity by fusing all the information that is used for the diagnosis of diseases. When this is transmitted in the common network, there is no issue of data spoofing, and when blockchain is used for authentication, the system security increases significantly [30]. For the data-loaded networks, the multi-view learning approach is used for analysing the network before transmission of information, where data fusion highly helps in the spoofing attacks [31]. The abnormal factors of Parkinson's disease are identified from the sensor data and are extracted for the purpose of multi-sensor fusion that helps in the detection and diagnosis of the disease at early stages [32].

2.5. Critical Observations and Research Gaps

Despite the rapid progress in IoMT data fusion, numerous key challenges remain unsolved. One crucial factor is computational competence, as several deep fusion models require considerable processing resources, limiting their suitability for real-time applications in resource-constrained IoMT devices. In terms of privacy protection, though privacy-enriched fusion methods referred to in [14] have been proposed, scalable and decentralised frameworks capable of assembling a complex healthcare system are still deficient. An additional hole lies in the domain-definite version, where recent fusion models are mainly widespread; the progress of dedicated architectures customised to medical sub-domains such as cardiology, neurology, or orthopaedics presents a capable path.

The subject of clarification is also leaves unexplored, through several mechanisms that deal with the interpretation of fusion data, a feature significant for both clinical trust and regulatory observance. Finally, integration with Edge AI appears as an important direction, as coupling edge computing with adaptive fusion could diminish latency and network dependence, and thereby attract sensitivity in IoMT healthcare systems.

The literature from [4, 6-24] provides an overview of classic fusion theory, signal-processing pipelines, attention-driven and transformer-based fusion, lightweight on-device approaches, and system-level privacy/security designs. Outstanding development consists of cross-attention and multi-scale transformers, wavelet-based preprocessing, and efficient fusion modules for wearables, all converging to enable precise, real-time, and personalised medical interventions. While AI-driven multimodal fusion and attention-based architectures control the latest advances, interpreting these advances into clinical practice needs investment in consistent standards, clear frameworks, privacy-utility trade-off studies, and rigorous real-world operation assessment. Dealing with these challenges together with computational efficiency will be crucial to unlocking the complete potential of IoMT-enabled innovative healthcare systems.

From the reviews, it is seen that the IoMT data fusion lacks cross-model dependency modelling, scalability issues, real-time systems inefficiency, and most of the work concentrates on the concatenation of features at the decision level, which overlooks the latent interactions at a deeper level.

Therefore, the proposed work encourages feature learning using the Variational Autoencoder for enabling cross-model data fusion using the Multi-Head Cross-attention mechanism for IoMT applications.

3. Data Fusion

Various techniques, such as compression methods, are available for data reduction in IoT networks, but these techniques are not efficient enough to accurately represent the patient data. In the IoMT system, the sensor data are of significant concern, as all patient information helps in disease diagnosis. So loss of information and reduction of data value lead to inaccuracy in the predictions. Therefore, to overcome the limitation of the data discrepancy from various single sources of the patient monitoring systems and to understand the condition of the patient in a reliable manner is necessary, which can be done using the data fusion method. Here, the homogeneous and heterogeneous data of multiple formats are processed to produce the fused data, which represents the entire dataset. Data fusion is the intelligent way of representing the system, more than aggregation or compression of the information.

In the IoT network, the data fusion is shown in Figure 1 is applied at various levels of the architecture, which has its pros and cons. Based on the level of fusion, the explanation is given below,

Low-Level Fusion: It is also known as the data level fusion or early fusion, where the most detailed and comprehensive information from the sensor level that provides a high signal-to-noise ratio, improves the data quality and enhances the completeness of the data. It involves the direct fusion of data from multiple sources before feature extraction. It has the disadvantages of high sensitivity for synchronisation alignment of sensors, intensive computation, and sensitivity to noise from various sources. The examples include image fusion, robotics, medical imaging, etc.

Mid-Level Fusion: In this type, the relevant information from multiple sources of data is considered for fusion, which is later used for analysis and classification purposes. Computation efficiency increases due to the analysis using relevant information alone. The efficiency of the fusion depends on how effectively the feature extraction takes place. This type finds significant application in autonomous vehicles for object detection and Biometrics.

High-Level Fusion: The fusion takes place at the decision level when the classification of the features is done, which are combined to produce a single decision. There is no synchronisation required between the raw data, which creates a highly reliable final decision by ensuring the independence of data sources. It is not suitable for weak classifiers as they produce highly correlated errors, and this fusion method is less detailed due to the reduction of context of the initial information. It is highly applicable in surveillance systems and medical diagnostics.

The data fusion in the IoMT system needs to address the characteristics like combined data from multiple sources should be free from errors, noises, and inconsistencies of its own sources that makes the output trustworthy, the fuion of information should take place in the faster rate that has to meet the real time capability of the IoMT system emergency situations, real time processing on the other hand helps in providing the critical health alert or timely decision making that requires the compuatationally efficient algorithm to give the faster results of the patient condition, in the growing exponential way of devices and senors in the IoT market, the data fusion method developed to be scalable that opt for fusing the information from the new sources even, the robust data fusion methods employs the continuous functioning capability even when the data imperfections, missing sensor values, sensor failures and errorneous readings, harmonization of data of various formats, resolutions and characteristics that should normalize align, and integrate.

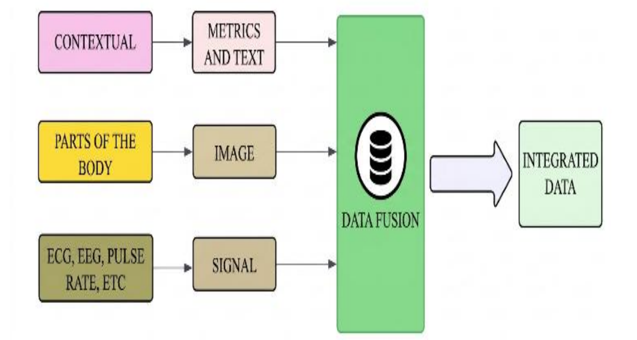

Fig. 1 Data fusion

Multimodal data, and the most crucial action of data fusion is to ensure strong security and privacy measures that maintain compliance with regulations like HIPAA and GDPR.

4. Proposed Work

The fusion of heterogeneous data in the IoMT systems to provide a single representation is the objective of this work, where the data fusion is exhibited using the variational autoencoder and cross-attention model that enables the single space representation of the raw data by latent representation and attention mechanism. The proposed work applies to text, numeric, context, image, and signal-based representations of information.

An IoT network from the patient side is used to collect the data, and the collected data are transmitted to the nearby node or the edge computing node. Edge computing is an efficient way to carry on the data fusion process, as fusion at the sensor level and final decision level is ineffective due to correlation issues.

In the proposed work, the data fusion is done at the mid-feature level, where the extraction of features is carried out using the Variational Autoencoder. The extracted features are provided with distinct attention by the cross-attention model for data fusion of the features. Figure 2 is the proposed architecture depicting the data fusion model considering the modal image, signal and numeric data-based inputs.

The working mechanism of the proposed architecture is explained in detail in the following steps to describe the data fusion process explicitly.

- The input modality of numeric data, signal or image data is obtained from the patient side of the IoT network.
- At the edge computing, the raw information is converted to digital (ADC), filtering, compression, and preprocessing by a normalisation process, etc, are carried out to obtain the actual information content data.

The information from the IoT network is separated based on the data type after the preprocessing step. For the proposed work, a multi-head attention-based cross-attention model [29] is used as it captures more relevant information across multiple criteria, facilitating data fusion.

- In the initial step, the latent dimensions are fed to the Query(Q), Key (K) and Value(V) neural network of the attention model.
- The Query is to obtain the target features to determine information from the latent 2, Key is the relevant information based on the Query parameter from latent1, and Value is the actual information content in the latent 2.

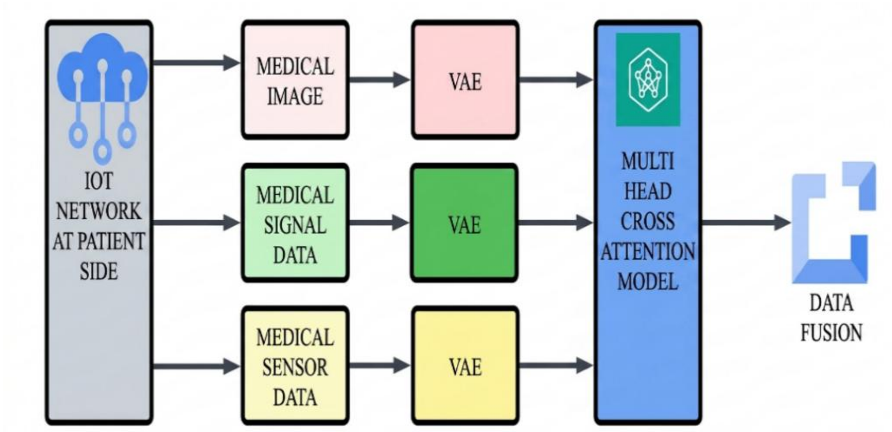

Fig. 2 Proposed framework

- Initially, in the attention heads, the query and key values interact to obtain the attention score, which is the actual representation of what relevant information is required in the latent space from the latent space 2.
- Then the attention score at the final attention head is computed by matrix multiplication of the above attention score and the value function, which produces the output that contains relevant information of latent two that is included in latent1.
- The output of the attention heads is concatenated to provide a single representation and is fed to the feed-forward neural network.
- In the FFNN, the final probabilities for the concatenated cross-attention model output are obtained.
- This is the actual data fusion output that represents the single representation of two different modalities.

Thus, by the above process, the data-fused output is obtained, which contains data not present in the raw data. This provides high data integrity of patient information using the proposed work.

$$X^{(m)} = \left\{X_i^{(1)}, X_i^{(2)}, \dots \dots . X_i^{(M)}\right\}_{n=1}^{N} \tag{1}$$

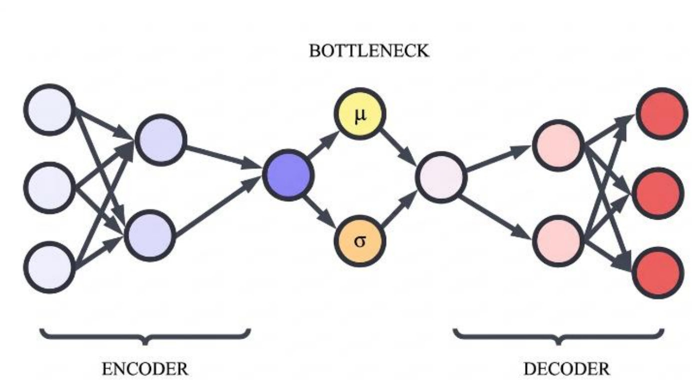

Fig. 3 Variational autoencoder architecture

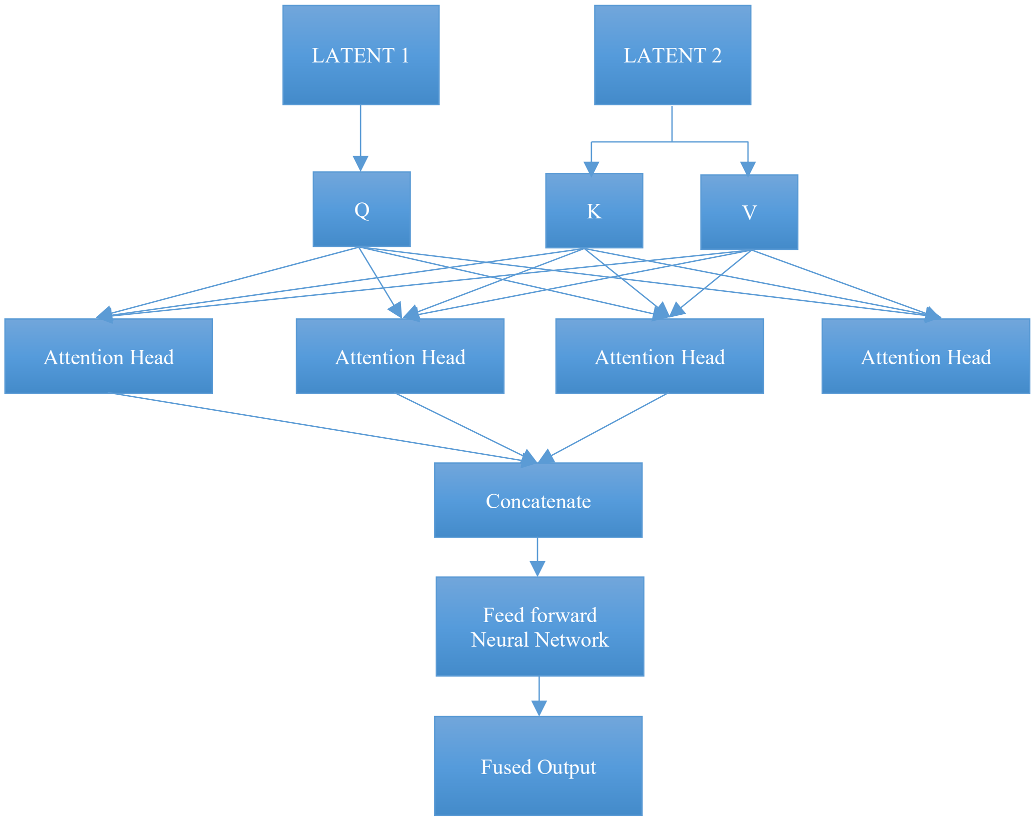

Fig. 4 MultiHeadCross attention model

In (1), N is the number of users that represent the 'M' modalities. Let us consider X as the preprocessed data; therefore, the X data is used for obtaining the variational autoencoder latent dimension given in Equation (2), where θm is the parametrization for the decoding process, z(M) is the latent representation,

$$logp_{\theta_m}(X^{(M)}) = log \int p_{\theta_m}(X^{(M)})|z^{(M)})p(z^{(M)}).dz^{(M)} \quad (2)$$

The encoder output of the Variational Autoencoder is given in Equation (3), where q(.) is the Variational posteriori function.

$$E_M = q_{\varphi_m}(z^{(m)}|X^{(m)}) \quad (3)$$

The initialised weights for the attention networks are WQ, WK, and WV, with dimensions of the key and query represented by dk, which is considered for every head. Using the weight matrix, the token matrix for Q, K and V is given by Equation (4),

$$Q^h = E_M.W_Q^h,\ K^h = E_M.W_K^h,\ V^h = E_M.W_V^h \quad (4)$$

Using these tokens, the final attention scores from the attention heads are given in Equation (5),

$$A^h = softmax\left(\frac{Q^h.(K^h)^T}{\sqrt{d_k}}\right).V^h \quad (5)$$

The concatenation of different head outputs is done by Equation (6),

$$A = [A^1, A^2, \dots \dots A^M]_{n=1}^N \quad (6)$$

The concatenated outputs are fed into the Feed Forward Neural Network that adds residual and layer normalisation, and are given in Equation (7) using the ReLU activation σ and weight for layer i given by Wi.

$$F = \sigma_i.A + W_i \quad (7)$$

Thus, the F is the data fused output of the multi-head cross attention model using the latent representation from the Variational Autoencoder.

Figure 5 is the data flow and process that takes place in the proposed data fusion model. The data from two different modalities are obtained and fed into different VAEs to get the latent space representation. The latent data are used to obtain the Q, K, and V parameters. Different attention scores are received from the four attention heads, which are consolidated, and the final fused output is obtained from the FFNN.

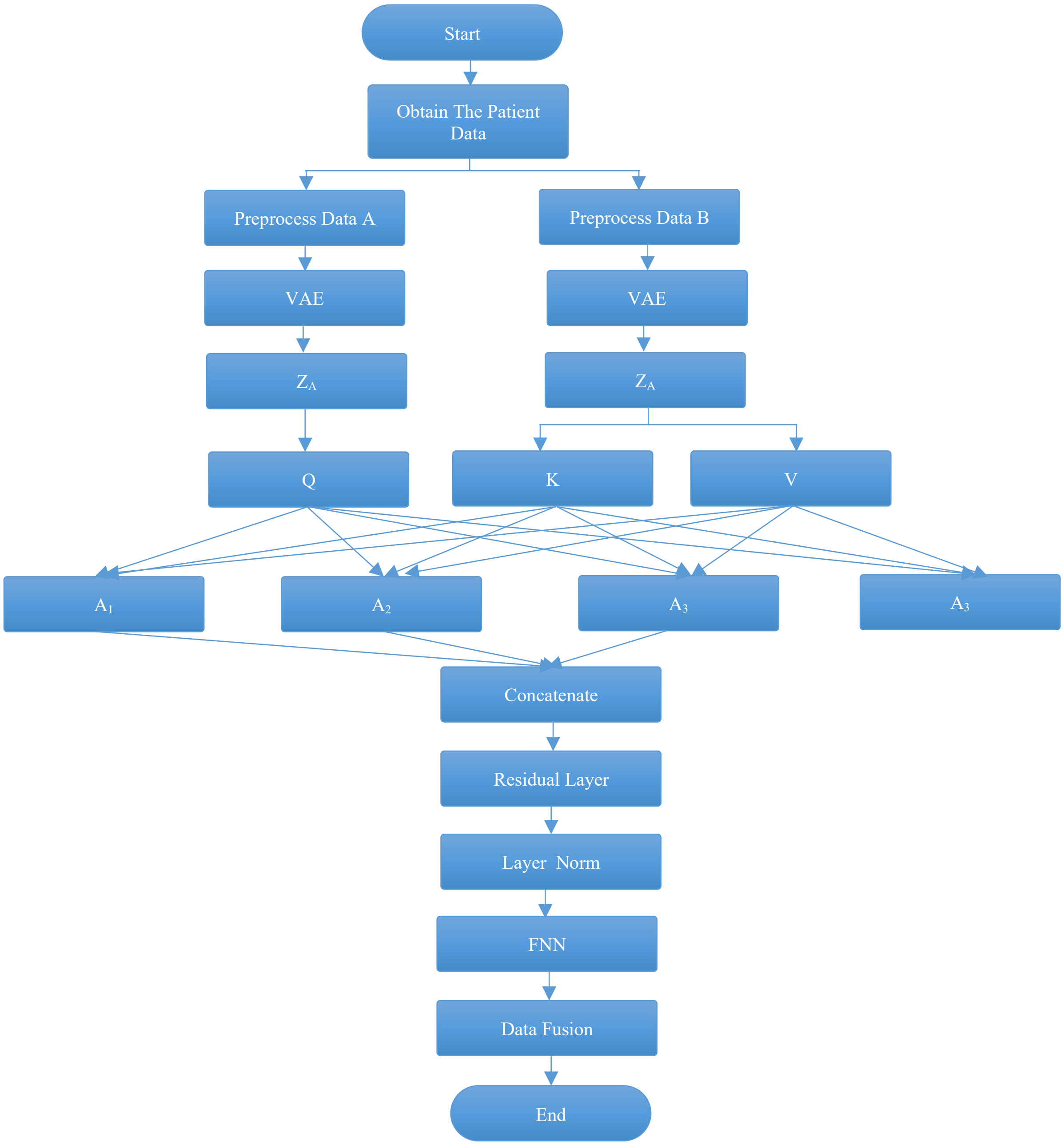

Fig. 5 Flow chart for the proposed data fusion model

Algorithm 1: Data Fusion using Cross-Attention of Latent Representation
1. Obtain the input multimodal data for N patients,   n {1, 2, …, N}   2. Data pre-processing using min-max scaling to obtain the normalised values.   3. The modality representation for each patient is given by, $X^{(m)} = \left\{X_i^{(1)}, X_i^{(2)}, \dots \dots . X_i^{(M)}\right\}_{n=1}^{N}$ 4. for i=1:n   5. for j=1:M   6. Obtain the latent representation of the modals using the encoder of the Variational Autoencoder by, $E_M = q_{\varphi_m}(z^{(m)} \mid X^{(m)})$

end
end
7. for i=1:n
8. Obtain the tokens,

$$Q^h = E_M.W_Q^h,\ K^h = E_M.W_k^h,\ V^h = E_M.W_V^h$$

9. Pass the Q, K, V into four different Attention heads.
10. Obtain the attention head score by,

$$A^h = softmax\left(\frac{Q^h.(K^h)^T}{\sqrt{d_k}}\right).V^h$$

11. Concatenate the attention scores for every single patient,

$$concat_{i=1}^n = \{A^1, A^2, \ldots \ldots A^h\}_{i=1}^n$$

12. Pass the concatenated output to the FFNN and obtain the production for n patients by

$$F_{i=1}^N = \left(\sigma_j.A + W_j\right)_{i=1}^n$$

end

5. Results and Discussions

The Variational Autoencoder-based Cross Attention model data fusion for IoMT applications framework is implemented in Python 3.11 using the Pytorch 2.2 deep learning library. The workstation utilised is NVIDIA RTX 4090 GPU (24 GB VRAM), Intel i9 CPU (3.6GHz, 16 cores) and 64GB RAM. Libraries like Scikit-learn are used for preprocessing, metrics calculation, and comparison with a baseline. For image transformations and feature extractions, TorchVision, NLTK, and Genism are used for text tokenisation and stopword removal. For producing the visual plots, Matplotlib and seaborn are used.

The proposed work is simulated using the parameters provided in Table 1, and the results are validated using the Digital Twin: EHR, Imaging, and IoT dataset from Kaggle. This dataset contains the demographics, medical history, and diagnosis patterns of the patient in the EHR file and a radiology-based imaging dataset that provides the spatial and anatomical state of the patient. In this dataset, the IoT-based data provides real-time monitoring actions, such as heart rate analysis, activity levels, and indicators based on mental health, with 20 records and 16 attributes. The EHR dataset contains the metadata like patient age, gender, medical history, diagnosis category, medication details and the clinical condition indicators.

The medical images in the dataset provide the details of the disease progression, structural deviations and changes in the physiological factors in a visual form. IoT sensor data acts as the real-time monitoring information in the proposed work that gives insights into the physiological signals.

Table 1. Simulation parameters

	PARAMETERS	**VALUES**
	VAE	
	DATA DIMENSIONS	20 FEATURES
	BATCHSIZE	64
	ENCODER LAYERS	2 FC
	LATENT DIMENSION	64
	ACTIVATION	RELU
	LEARNING RATE	0.0001
	OPTIMIZER	ADAM
	LOSS FUNCTION	KL DIVERGENCE
CROSS ATTENTION MODEL		
NO OF ATTENTION HEADS	4	
ATTENTION DIMENSION	64	
Q/K/V PER HEAD	16	
DRPOUT RATE	0.1	
FUSION OUTPUT DIMENSION	128	
NO OF ATTENTION HEADS	4	
ATTENTION DIMENSION	64	
	FFNN	

	HIDDEN LAYERS	2 FC
	HIDDEN UNITS	[128, 64]
	OUTPUT LAYER	SOFTMAX
	LEARNING RATE	0.0001
	OPTIMIZER	ADAM
	HIDDEN LAYERS	2 FC
	HIDDEN LAYERS	2 FC
	HIDDEN UNITS	[128, 64]

In the preprocessing of the images, it is resized to 224×224 pixels, normalization with Imagenet mean and variance. In Figure 6, the VAE latent representation for the three modalities is shown, considering the variance of the X-axis with 15.39% and the Y-axis with 13.47%, which represent the modalities of the HER dataset, Image features, and IoT sensor data. The latent dimensions spread out evenly, but some features overlap, leading to the discrimination of data that helps capture the variability in the data.

Figure 7 shows the features that significantly contribute to the data fusion, where the image contribution is higher than 0.20, which acts as the most influential element in the fusion process. It is followed by the age factor of 0.18, which contributes to the demographic factor, and the heart rate of 0.15 for the physiological aspects. In this way, the data fusion process explicitly reveals its potential by considering important features for patient condition diagnosis.

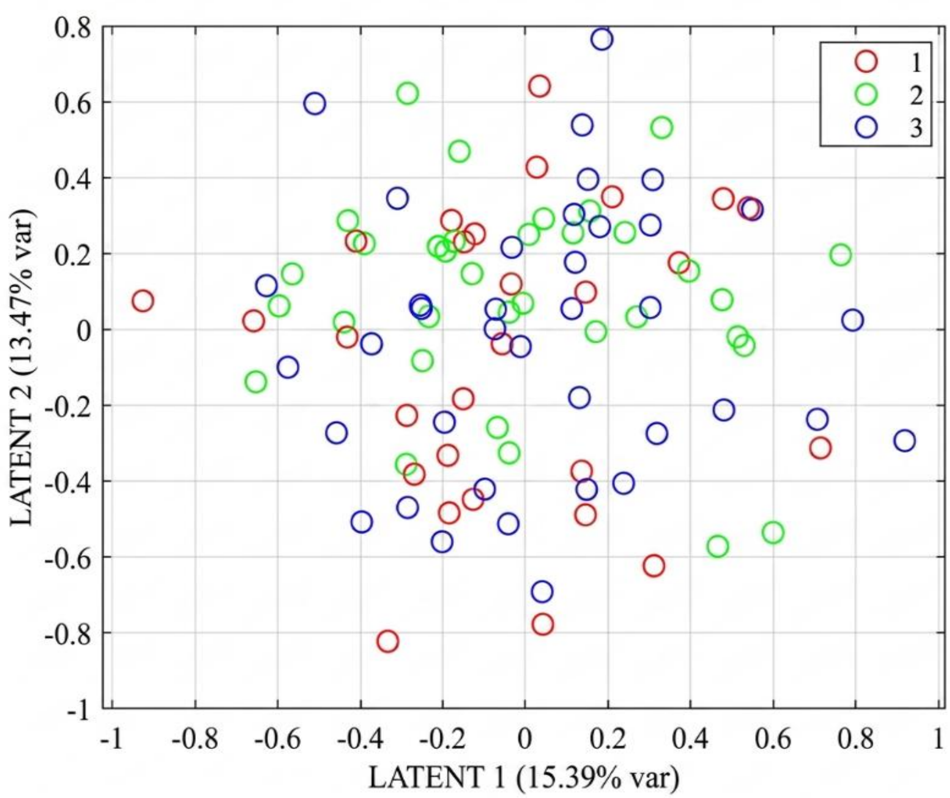

Fig. 6 VAE latent representation for three modals, 1 for HER, 2 for medical image features and 3 for IoT sensor features

Figure 8 is the VAE reconstruction loss curve using the KL divergence theorem for the mean and variance distribution in the latent representation. To study the performance of the VAE, the decoder part is separately constructed, and the loss curve is plotted. In the epochs 1 to 8, there is a steady drop in the curve, indicating the effective learning process, and it saturates after that, where the learning process convergence starts after 10 epochs. Major patterns are learned in the earlier stages, followed by the specific pattern study near the convergence scale. Thus, the reconstruction quality improves over time effectively by learning the latent representation.

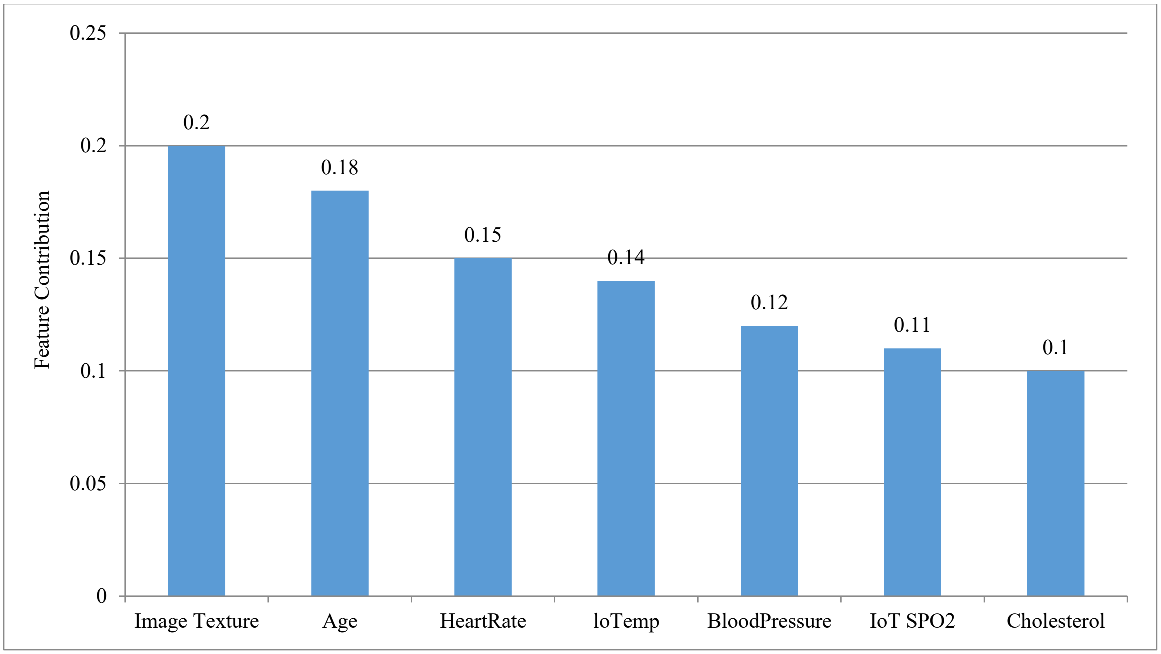

Fig. 7 Samples of feature contributions in data fusion using VAE

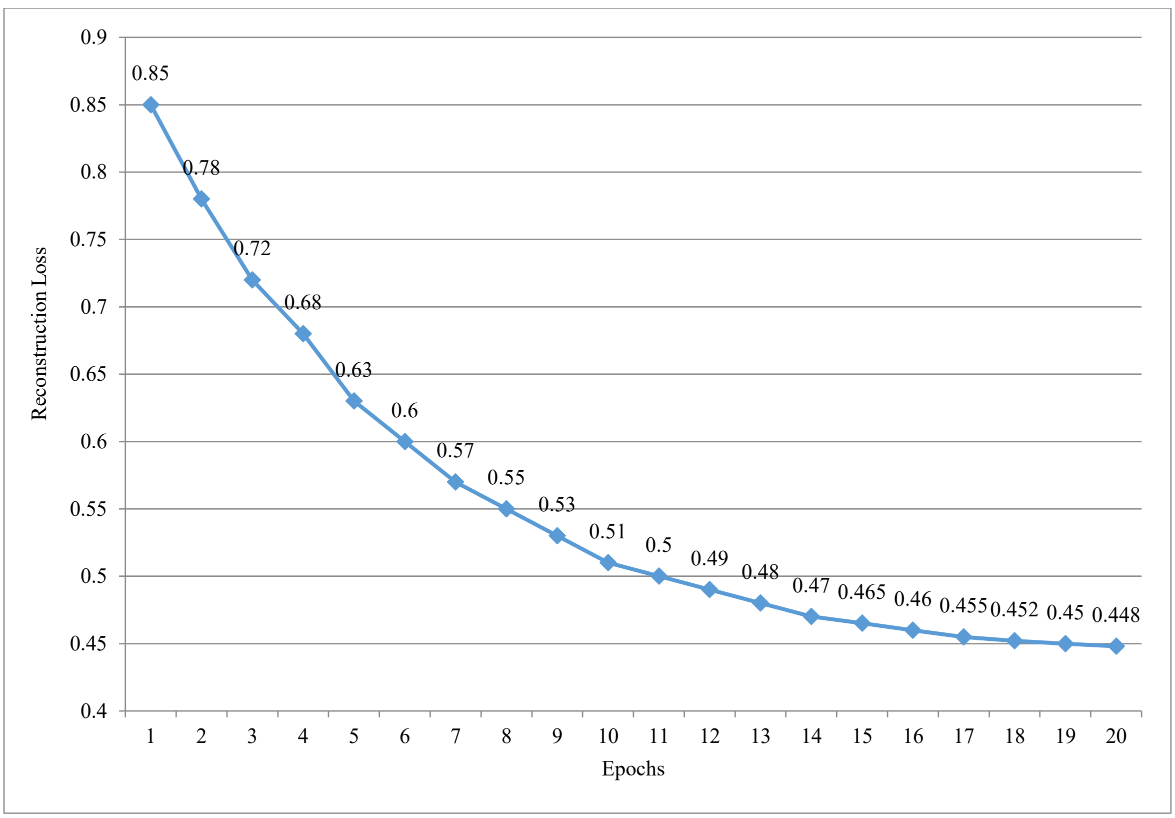

Fig. 8 Reconstruction loss curve considering the labelled dataset for VAE evaluation

Figure 9 shows the attention scores of the single attention head to showcase the feature importance given by the query and key for different data formats. The highest attention score is at 0.5 for the query feature of 2 and the key feature of 5. There is no correlation between the identical features that are seen, as research has shown that the attention mechanism is quite directional. Query and Key are the main parameters contributing to the cross-attention model that indicate the clinically relevant multimodal relationship.

Figure 10 is the prediction status using the risk scores for single modal using EHR, Image only and IoT sensor data with the proposed multimodal fusion data. LSTM is used in place of FFNN to study the real-time monitoring system, where the proposed work outperforms other single modality predictions. The IoT sensor data serves as the basis for improvement, but the proposed work outperforms it by 2.2%.

Table 2 is the evaluation metrics for the proposed work, and the results are compared with the single modal prediction data basis. The accuracy of the proposed work outperforms others by 10% of IoT sensor data, 6% by Imaging data and 8% in terms of EHR data. AUC is high at 97.3%, which shows the trade-off between the actual positive and false negative values.

Table 2. Evaluation metrics

PARAMETERS	EHR	IMAGING	IOT	PROPOSED WORK
ACCURACY	86.5%	88.2%	84.7%	94.6%
PRECISION	85.9%	87.5%	83.8%	94.1%
RECALL	84.7%	86.4%	82.5%	93.5%
F1	85.3%	86.9%	83.1%	93.8%
AUC	90.1%	91.4%	88.6%	97.3%

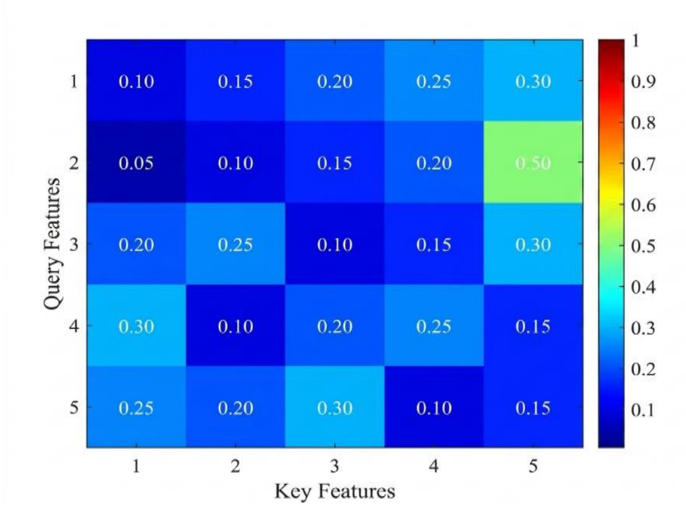

Fig. 9 Attention scores for query and key dimensions

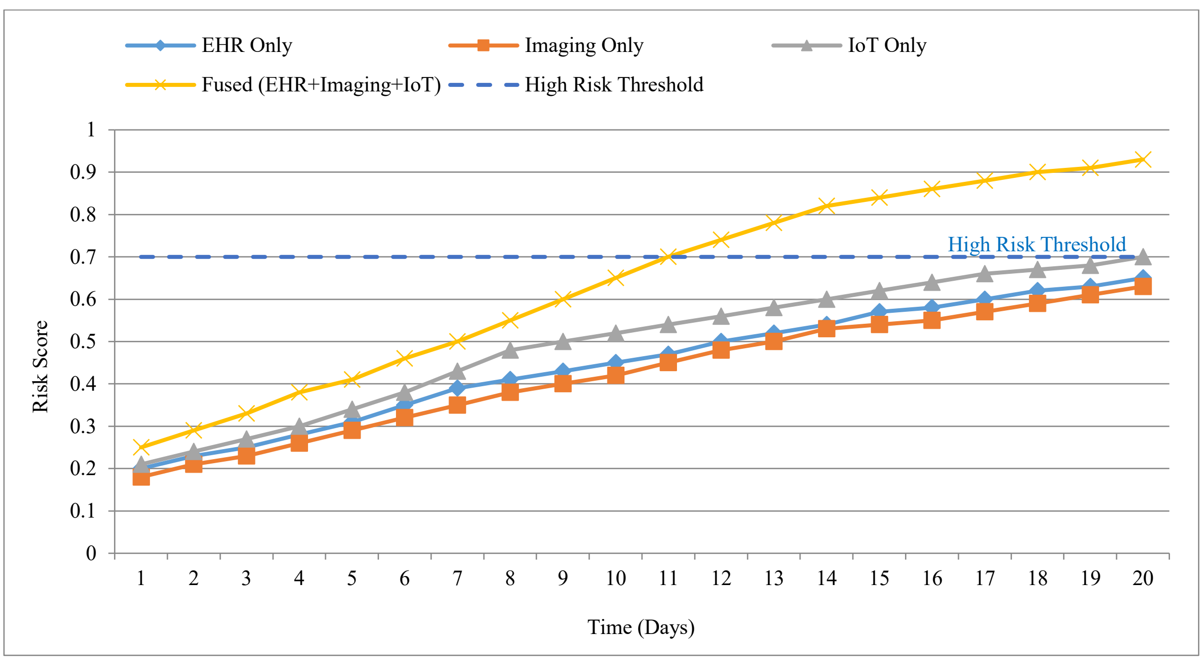

Fig. 10 Comparison of prediction of health status

Table 3. Comparison table

REF	MODEL	ACCURACY
[7]	Multi-Modal Sensor Fusion for Posture Prediction	82-89%
[10]	Sensor Data Fusion	89%
[12]	Wavelet Entropy-based Fusion	91%
[16]	Self Attention Fusion	90-92%
Proposed Methodology		**94.6%**

Table 3 discusses the accuracy values of the baseline models of the proposed work, where the early fusion, late fusion, and attention-based fusion in the surveyed literature. In the proposed work, the abstracted information that is obtained using the VAE and global attention paid to the features among different data helps in obtaining the maximum accuracy.

6. Proposed Work System Considerations and Deployment Aspects

The proposed work's interpretability is provided in Figure 9, which shows how the query and key value interact to obtain the important information that is present in the EHR data and the image data. This helps in visualising the important feature responsible for the disease diagnosis, which is multiplied by the conventional important value data. This provides the feature fusion based on global attention with the abstracted data. This proposed system is applicable to function in the edge computing devices of the IoMT system, which provides a unified representation of the diagnosis along with a reduction in the network load. The unified information produces raw data security by not sharing the original information in the network.

7. Conclusion

Data fusion plays a prominent role in the data integration process and helps to represent a considerable amount of data into reduced dimensionality. IoMT systems cannot handle the vast amount of data and heterogeneous data types for each patient to accurately predict the condition. Based on this objective, the proposed work used a latent representation of input heterogeneous modality using a Variational Autoencoder and fused it using the Multi-Head Cross-Attention Model. The proposed work is validated using the EHR, Image, and IoT sensor dataset from Kaggle, achieving an accuracy of 94.6%, which outperforms the predictions by single modal data by 8%. The AUC of the proposed work is 97.3%, indicating the proposed work's efficacy in classification processes. The proposed work suffers from computational complexity due to the increased number of neural networks in the VAE block, Q, K, and V blocks of the attention mechanism.

The proposed work applies to data fusion in various applications and scalable datasets. The extension of the proposed work is to use it in the decentralised system for the IoMT system to ensure privacy and enable multiple hospitals for improved generalisation processes.

Conflicts of Interest

The Authors ensure no conflict of interest in this publication.

Funding Statement

This work is not supported by any Funding Organizations.

References

[1] Ali Ghubaish et al., "Recent Advances in the Internet-of-Medical-Things (IoMT) Systems Security," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8707-8718, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Tahir Abbas et al., "IoMT-based Healthcare Systems: A Review," *Computer Systems Science and Engineering*, vol. 48, no. 4, pp. 871-895, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Amirhossein Adavoudi Jolfaei, Seyed Farhad Aghili, and Dave Singelee, "A Survey on Blockchain-based IoMT Systems: Towards Scalability," *IEEE Access*, vol. 9, pp. 148948-148975, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Muhammad Shafiq et al., "Advances in IoMT for Healthcare Systems," *Sensors*, vol. 24, no. 1, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Imran Ahmed, Eulalia Balestrieri, and Francesco Lamonaca, "IoMT-based Biomedical Measurement Systems for Healthcare Monitoring: A Review," *Acta IMEKO*, vol. 10, no. 2, pp. 174-184, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Shams Forruque Ahmed et al., "Insights into Internet of Medical Things (IoMT): Data Fusion, Security Issues and Potential Solutions," *Information Fusion*, vol. 102, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Taher M. Ghazal et al., "IoMT-Enabled Fusion-based Model to Predict Posture for Smart Healthcare Systems," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 2579-2597, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mian Ahmad Jan et al., "Lightweight and Smart Data Fusion Approaches for Wearable Devices of the Internet of Medical Things," *Information Fusion*, vol. 103, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Hemant Ghayvat et al., "Revolutionising Healthcare: IoMT-Enabled Digital Enhancement via Multimodal ADL Data Fusion," *Information Fusion*, vol. 111, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Fan Yang et al., "Internet-of-Things-Enabled Data Fusion Method for Sleep Healthcare Applications," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15892-15905, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xuhui Ding et al., "A Feasibility Study of Multi-Mode Intelligent Fusion Medical Data Transmission Technology of Industrial Internet of Things Combined with Medical Internet of Things," *Internet of Things*, vol. 21, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[12] MD. Mobin Akhtar et al., "IoMT-based Smart Healthcare Monitoring System using Adaptive Wavelet Entropy Deep Feature Fusion and Improved RNN," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 17353-17390, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[13] Veselin Atanasov, and Yordan Sivkov, "Data Fusion for IoMT in Shipping," *2020 21st International Symposium on Electrical Apparatus and Technologies (SIELA)*, Bourgas, Bulgaria, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[14] Hui Lin et al., "Privacy-Enhanced Data Fusion for COVID-19 Applications in Intelligent Internet of Medical Things," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15683-15693, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[15] Fei Zhao, Chengcui Zhang, and Baocheng Geng, "Deep Multimodal Data Fusion," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1-36, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[16] Xuezhu Lin et al., "Multi-Sensor Data Fusion Method based on Self-Attention Mechanism," *Applied Sciences*, vol. 13, no. 21, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[17] Silvana M. Azcarate et al., "Data Handling in Data Fusion: Methodologies and Applications," *TrAC Trends in Analytical Chemistry*, vol. 143, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[18] Federico Castanedo, "A Review of Data Fusion Techniques," *The Scientific World Journal*, vol. 2013, no. 1, pp. 1-19, 2013. [CrossRef] [Google Scholar] [Publisher Link]
[19] Yunhao Gao et al., "Cross-Scale Mixing Attention for Multisource Remote Sensing Data Fusion and Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[20] Zilong Huang et al., "CCNET: Criss-Cross Attention for Semantic Segmentation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 603-612. 2019. [Google Scholar] [Publisher Link]
[21] Chun-Fu Richard Chen, Quanfu Fan, and Rameswar Panda, "CrossVIT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 347-356, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[22] Hezheng Lin et al., "Cat: Cross Attention in Vision Transformer," *2022 IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[23] Mozhdeh Gheini, Xiang Ren, and Jonathan May, "Cross-Attention is All You Need: Adapting Pretrained Transformers for Machine Translation," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics,
Online and Punta Cana, Dominican Republic, pp. 1754-1765, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[24] Hui Li, and Xiao-Jun Wu, "CrossFuse: A Novel Cross Attention Mechanism based Infrared and Visible Image Fusion Approach," *Information Fusion*, vol. 103, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[25] Wenjuan Han, Ge Wang, and Kewei Tu, "Latent Variable Autoencoder," *IEEE Access*, vol. 7, pp. 48514-48523, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[26] Junhai Zhai et al., "Autoencoder and its Various Variants," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, pp. 415-419, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[27] Fei Yang et al., "Variable Rate Deep Image Compression with Modulated Autoencoder," *IEEE Signal Processing Letters*, vol. 27, pp. 331-335, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[28] Jinwon An, and Sungzoon Cho, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability," *Special Lecture on IE*, vol. 2, no. 1, pp. 1-18, 2015. [Google Scholar]
[29] Lucas Pinheiro Cinelli et al., *Variational Autoencoder*, Variational Methods for Machine Learning with Applications to Deep Networks, Springer, Cham, pp. 111-149, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[30] Abdullah Ayub Khan et al., "Blockchain-Enabled Secure Internet of Medical Things (IoMT) Architecture for Multi-Modal Data Fusion in Precision Cancer Diagnosis and Continuous Monitoring," *Journal of Cloud Computing*, vol. 14, no. 1, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[31] Sekione Reward Jeremiah et al., "Multi-View Learning and Model Fusion Framework for Threat Detection in Multi-Protocol IoMT Networks," *Information Fusion*, vol. 125, 2026. [CrossRef] [Google Scholar] [Publisher Link]
[32] Tongyue He et al., "Enhanced Detection of Early Parkinson's Disease through Multi-Sensor Fusion on Smartphone-based IoMT Platforms," *Information Fusion*, vol. 117, 2025. [CrossRef] [Google Scholar] [Publisher Link]