RESEARCH PAPER

Multimodal Machine Learning for Biomedical Data Integration in Medicine

Muath Khawaji[1]

[1] Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia. Email: m.khawaji@jazanu.edu.sa

Received: 20th Feb, 2026 | **Revised:** 4th Mar, 2026 | **Accepted:** 25th Mar, 2026 | **Available Online:** 10th Apr, 2026

ABSTRACT

The growing accessibility of nonhomogeneous biomedical information has opened up new prospects in the field of developing healthcare by means of intelligent integration of data. Multimodal machine learning offers an effective platform to integrate heterogeneous sources of data, such as medical images, genomic data, electronic health records, and wearable sensor data, to provide a better experience in clinical decision-making. The paper provides a review of multimodal data integration techniques in medicine, including preprocessing techniques, representation learning, fusion techniques and system architectures. There are many types of fusion, including early, late, and hybrid fusion, which are explained to provide an overview of their merits and demerits in biomedical practice. A system architecture that is layered is recommended to be used in order to achieve effective data acquisition, preprocessing, learning, and decision support. Multimodal methods are proven to increase the diagnostic accuracy and predictive performance in clinical practice in such areas as oncology, cardiovascular diseases and neurological disorders. They discuss evaluation and benchmarking methodology, as it is required to ensure that it is robust, scalable, and applicable in the real world. The issues of the heterogeneity of the data, the absence of modalities and the problems of interpretability and privacy are also regarded as the major challenges. The findings have revealed that multimodal machine learning is much better than the traditional unimodal algorithms and has numerous possibilities to deliver precision medicine and intelligent healthcare systems.

Keywords: Multimodal Machine Learning, Biomedical Data Integration, Medical Imaging, Electronic Health Records, Genomic Data.

How to cite this article: Khawaji M. Multimodal Machine Learning for Biomedical Data Integration in Medicine. Int J Drug Deliv Technol. 2026;16(31s):384. DOI: 10.25258/ijddt.16.31s.46

Source of support: Nil.

Conflict of interest: The author declares no conflict of interest.

I. Introduction

The rapid digitalization of healthcare systems has brought about a radical growth in the quantity and diversity of biomedical data being generated both in clinical and research environments. The trend of modern medical practice is to use heterogeneous medical data sources, such as medical imaging modalities, such as magnetic resonance imaging (MRI) and computed tomography (CT), genomic and proteomic data, electronic health records (EHRs), and continuous physiological measurements provided by wearable devices. Such varied streams of data give different insights into the health of a patient, but due to their scattered and unorganized structure, it is very difficult to integrate and analyze them effectively [1]. The need to decode these kinds of multimodal data into uniform and useful information has been one of the objectives of improving precision medicine and improving clinical decision-making. Multimodal machine learning is a powerful paradigm that has been designed to address these problems by enabling the analysis of various modalities of data. In contrast to conventional machine learning models, which use one data source, multimodal models are developed to reflect intricate interrelationships and latent associations among heterogeneous data. The systems can learn high-level representations of each modality and jointly schedule them into single predictive models using the newest learning architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models [2]. It would be applicable in the medical practice where imaging, clinical and molecular information may be integrated to immensely contribute to the diagnostic accuracy, prognostic assessment and treatment strategies.

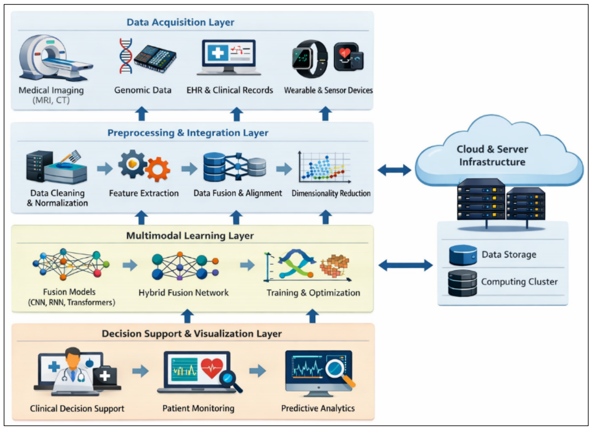

Figure 1. Block Schematic for Multimodal Biomedical Data Integration & Clinical Decision Support

Multimodal data combination has been quite a promising area in many clinical domains. In cancer, e.g. the combination of radiological data with genomic results might result in a more precise characterization of tumors and selection of individual treatment options as illustrated in figure 1. Electrocardiogram (ECG) signals combined with the cardiovascular medicine history and imaging of patients are useful in detecting complicated heart diseases. Similarly, in neurology, neuroimaging data with behavioral and electrophysiological data will aid in an early diagnosis of neurodegenerative disorders [3], [4]. Such applications indicate the revolutionary aspect of multimodal learning in delivering holistic and informational healthcare solutions. The problems of data heterogeneity, missing or incomplete modalities, high dimensionality, and different data quality make it complicated to design strong learning models. Besides, questions of data privacy, data security, and data interpretability remain the key challenges to the widespread adoption in clinical practice [5]. To deal with these issues, scalable architectures, effective fusion methods and clear models which can offer explainable information to medical practitioners need to be develope.

II. Biomedical Data Modalities and Characteristics

Modern medicine biomedical data has numerous sources and each of them is characterized by another aspect of human physiology, pathology and clinical history. Medical imaging data, such as magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, X-ray, and positron emission tomography (PET) gives detailed spatial and structural details of tissues and organs [6]. These imaging modalities are all essential in the detection of diseases, staging and monitoring of treatment. Their rich visual detail makes them very amenable to deep learning-based pattern recognition, but are also computationally demanding and frequently need significant preprocessing, such as denoising, segmentation, and spatial normalization. [7] The incompatibility of acquisition protocols and scanner settings and image resolution also adds to the difficulty of directly using imaging data across institutions and patient groups. Another important modality of biomedical analytics is genomic and molecular data. The information derived through DNA sequencing [8], RNA expression profiling, proteomics and metabolomics offers an insight into the biological events taken part in the disease progression and responses to the therapy. This kind of data is very dimensional in that it could have thousands of variables and yet have only a few numbers of patients' samples [9].

This imbalance is a severe risk of overfitting machine learning frameworks and necessitates advanced feature selection, dimensionality reduction and representation learning procedures [10]. Precision medicine depends especially on molecular data, as individualized biological signatures can be used to inform personalized medicine. However, this modality is difficult to integrate with other clinical data because of the scale disparity, scarcity and the need to interpret it in a domain-specific manner [11] [12]. Electronic health records (EHRs) add longitudinal and contextual information concerning patient care. They typically include demographic data, diagnosis codes, laboratory tests, history of medications, doctor notes, and procedure history. EHR data is useful as it is a representation of clinical practice and patient course in the real world. At the same time, it is highly heterogeneous, an assortment of the structured fields with the unstructured narrative text [13], [14]. The lack of values, bad coding conventions and fragmented record systems are major hindrances to successful computational analysis. The methods of natural language processing, and time modeling then are likely to be required regularly to derive meaningful representations on EHR repositories [15]. These modalities of time-series prove to be of great use in the management of chronic diseases, rehabilitation, and remote patient monitoring as outlined in table 1 data.

Table 1: Summary of Biomedical Data Modalities and Key Characteristics

Modality	Data Type	Key Features	Challenges
Medical Imaging (MRI, CT) [16]	Spatial / Visual	High-resolution structural information	High computation, variability
Genomic Data [17]	High-dimensional	Molecular-level insights, precision medicine	Data sparsity, overfitting risk
EHR (Clini [18]cal Records)	Structured + Text	Longitudinal patient history	Missing data, inconsistency
Wearable/Sensor Data	Time-series	Real-time monitoring, continuous signals	Noise, irregular sampling

III. Multimodal Data Preprocessing and Representation Learning

Multimodal biomedical data can only be effectively used with strong preprocessing and representation learning approaches that are capable of converting heterogeneous raw data into structured, informative and machine readable forms. Since data modalities are diverse, such as high-resolution imaging and high-dimensional genome profiles, unstructured clinical text and time-series physiological signals, preprocessing is an important aspect of maintaining data quality, consistency, and interoperability across modalities. The data normalization and data cleaning is the starting point of the preprocessing stage. Biomedical data is usually noisy, contains some gaps, and has inconsistencies and artifacts due to error in measurements, device constraints, or human input variability. The data that is obtained after the imaging process is usually subjected to denoising, contrast, and segmentation in order to extract areas of interest. Examples of normalization methods necessary to make comparisons between experiments in genomics data include a log transformation and a batch effect correction. EHRs require structured parsing, standardization of coding (e.g. ICD codes) and natural language processing algorithms to derive meaningful information out of clinical stories. Likewise, wearable sensor data have to be filtered to eliminate motion artifacts and resampled to overcome irregular time intervals. Convolutional neural networks (CNNs) are popular in imaging modalities to learn hierarchical spatial features, textures, shapes, and anatomical patterns, in an automatic manner. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have shown usefulness in time-series data models of the temporal dependencies and sequences, e.g. ECG and EEG systems. Transformer-based architectures, like Bidirectional Encoder Representations from Transformers (BERT), support semantic interpretation and contextual encoding of unstructured text in the context of clinical text. High dimensionality genomic data, typically, needs dimensionality reduction methods like the principal component analysis (PCA) or autoencoders to obtain small yet informative genomic representations as shown in figure 2. The multimodal systems representation learning aims at matching and uniting features of various types of data in a common latent space.

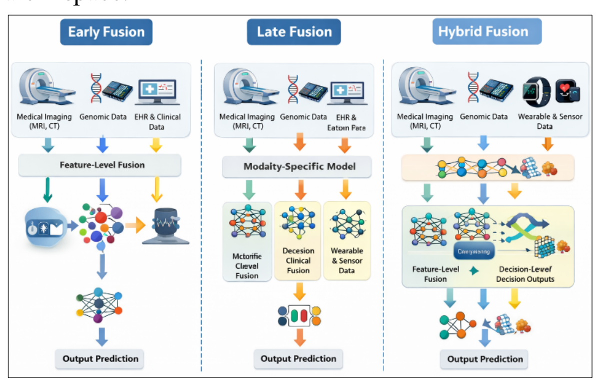

Figure 2. Early, late, and hybrid fusion strategies for multimodal biomedical data integration.

The other essential issue of multimodal representation learning is the ability to deal with the missing or incomplete modalities. In a real-life clinical situation, it is typical that some types of data may not be available to all patients. Modes (e.g., modality imputation, cross-modal learning, generative models such as variational autoencoders and generative adversarial networks) are applied to a missing information to estimate it and ensure the model remains robust. Also, data alignment methods are needed to align modalities of various temporal or spatial resolutions, which should be coherent. Other considerations in preprocessing and representation learning include scalability and computational efficiency. Large-scale biomedical datasets are typically processed with the help of distributed processing frameworks and cloud-based infrastructures. The emergence of effective pipelines integrating preprocessing, feature extraction and representation learning into a single workflow make multimodal systems more feasible in clinical practice. An appropriate preprocessing and representation strategy eventually lays the groundwork to proper multimodal fusion and predictive modeling as are further elaborated in subsequent subsections.

IV. Multimodal Fusion Strategies and Learning Architectures

The fundamental principles of integrated biomedical learning systems are multimodal fusion strategies, which allow using heterogeneous sources of data to build a single predictive framework. Fusion strategy has a profound impact on model performance, interpretability and computational efficiency. Multimodal fusion methods can be broadly divided into early fusion, late fusion and hybrid fusion, with each having different benefits based on the type of data and the clinical process. Early fusion, also known as feature-level fusion, is fusion that combines multiple modalities of data at the input stage by concatenating or combining the feature representations of these modalities into a single unified feature-vector. By this method, the model can learn the interactions of features between modalities, complex correlations among imaging, genomic and clinical data. Early fusion is especially effective in situations where modalities are complementary and share complementary information. Nevertheless, it also is susceptible to absent data and variation in scale of features, and can become a victim of the curse of dimensionality when dealing with high-dimensional biomedical datasets.

Table 2. Comparison of multimodal fusion strategies in biomedical applications

Fusion Strategy	Fusion Level	Key Strength	Limitation
Early Fusion	Feature-level	Captures cross-modal feature relations	High dimensionality, sensitive to missing data
Late Fusion	Decision-level	Modular, handles missing modalities	Limited cross-modal interaction
Hybrid Fusion	Feature + Decision	Combines strengths of both approaches	Higher complexity, computational cost

The decision-level or late fusion takes a different approach and works at the decision-level by independently processing the different modalities with the use of different models and then fuses their results at the decision-level. This is more suitable to modularity and robustness since every modality-specific model can be optimized separately as presented in table 2 data. It is especially helpful in the case when modalities are loosely coupled or some modalities are not available. Prediction aggregation methods like majority voting, weighted averaging, or meta-classifiers are popular ways of aggregating predictions. Although flexible, late fusion might not be able to capture rich cross-modal relationships, which would restrict its capability to fully harness the complementary quality of multimodal data. The hybrid fusion techniques seek to address the constraints of early and late fusion and combine modalities at different levels of the learning pipeline. These methods incorporate feature-level and decision-level fusion, and frequently based on intermediate representations learned by deep neural networks. Hybrid models can represent low-level feature interactions, high-level semantic relationships and thus can be used to address complex biomedical tasks. Often attention mechanisms and gating networks are added to dynamically weight the contribution of each modality according to its relevance to the task. A breakthrough in deep learning has made multimodal fusion architecture performance much more effective. Convolutional neural networks (CNNs) are popular to extract spatial information of imaging data whereas recurrent neural networks (RNNs) and long short-term memory (LSTM) networks represent temporal dependencies of physiological signals. Transformer-based architectures have been of recent interest as they can establish long-range dependencies and cross-modal interactions via self-attention. Multimodal transformers, especially, allow collaborative learning of the different types of data through the alignment of the representations in a common latent space.

V. System Architecture for Biomedical Data Integration

An effective system architecture is needed to facilitate effective integration, processing and analysis of multimodal biomedical data in the contemporary healthcare setting. The architecture proposed is based on a layered design, which guarantees the modularity, scalability and interoperability of heterogeneous data sources and computational elements. Such a structured design enables a smooth flow of data between acquisition and decision-making, enabling real-time clinical applications, and analytics at scale. The architecture starts with the data acquisition layer that is charged with the responsibility of acquiring biomedical data in various sources. Such sources are medical imaging systems that produce MRI, CT and X-ray data, genomic sequencing systems that produce high-dimensional molecular data, electronic health record systems that store structured and unstructured clinical data and wearable devices that record continuous physiological measurements, including heart rate, oxygen levels and activity patterns. The use of standardized communication protocols and healthcare data formats are used to provide interoperability and safe transmission of data as illustrated in figure 3. This layer serves as the interface of the system, at which the raw multimodal data is received and processed to be further processed.

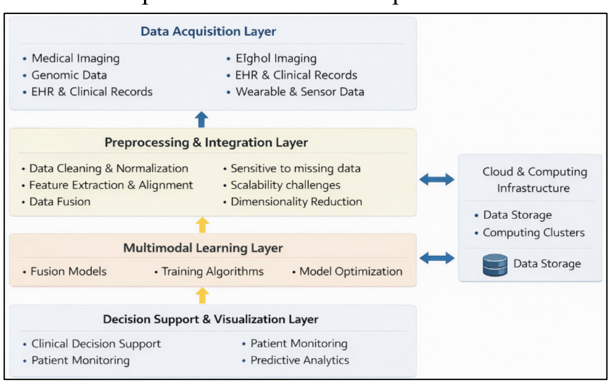

Figure 3. Biomedical Data Integration System Architecture

The second layer is the data preprocessing and integration layer which converts the raw data into a standard and analyzable format. The essential tasks executed by this layer include cleaning up the data, normalizing data, extracting features, and intermodal alignment. Imaging data are segmented and enhanced and genomic data is normalized and dimensionality reduced. EHRs have clinical text that is processed using natural language processing to obtain structured features. The data of time-series sensors is filtered and synchronized with the time to guarantee the temporal coherence. Fusion of data at this level can be either at feature level or embedding alignment to form a single representation. This layer makes sure that there is harmonization of heterogeneous data prior to entering the learning phase. The system analytical core is made up of the multimodal learning layer. It uses the state-of-the-art machine learning and deep learning algorithms to derive patterns and trends using integrated data. The spatial feature extraction of imaging data is performed with convolutional neural networks, whereas the temporal and contextual dependencies are learnt with recurrent and transformer-based models. This layer has multimodal fusion mechanisms, such as early, late, and hybrid strategies, that are used to effectively combine features and predictions. The representation learning techniques and attention mechanisms are used to dynamically balance the significance of both modalities allowing the system to concentrate on the most relevant information to complete a particular clinical task. The cloud and computing infrastructure supports these layers and offers the computing resources to work with large-scale biomedical datasets.

VI. Clinical Applications and Case Studies

Multimodal machine learning has shown great promise to revolutionize clinical practice by providing more precise, personalized and data-driven solutions to health care. The combination of heterogeneous sources of biomedical data enables clinicians to go beyond the individual analysis of each of the modalities and instead use complementary data to make better diagnosis, prognosis, and treatment plans. In this section, the main clinical applications and typical case scenarios in which multimodal approaches have demonstrated practical advantages are pointed out. Multimodal learning is a vital aspect in the diagnosis of cancer and effective treatment in oncology. Radiological imaging, the use of histopathological data, and genomic profiles allow a complete characterization of tumors. An example is that incorporation of features of MRI or CT imaging with features of gene expression could be used to better classify and stage tumors. Multimode models are able to find out subtle correlations between imaging phenotypes and molecular signatures, with the help of which individual therapy can be selected. Breast and lung cancer case studies have also shown multimodal systems to be better than unimodal systems in predicting patient survival rate and response to treatment, thus enhancing clinical outcomes.

Table 3. Multimodal machine learning applications across clinical domains and their outcomes

Application Domain	Modalities Used	Key Outcomes	Performance Impact
Oncology (Cancer Diagnosis)	MRI/CT + Genomic Data + Histopathology	Improved tumor classification and therapy selection	Higher diagnostic accuracy, better survival prediction
Cardiovascular Diseases	ECG + Echocardiography + EHR + Wearables	Early detection of arrhythmias and heart conditions	Reduced hospitalization, improved monitoring
Neurological Disorders	fMRI + EEG + Clinical Assessments	Early detection of Alzheimer's and Parkinson's	Increased classification accuracy
Critical Care (ICU / Sepsis)	Vital Signs + Lab Reports + EHR + Imaging	Real-time patient monitoring and risk prediction	Reduced mortality, faster clinical decisions
Remote Patient Monitoring	Wearable Sensors + EHR + Mobile Health Data	Continuous health tracking and anomaly detection	Improved preventive care and patient engagement

Another area where multimodal integration has been very effective is in the management of cardiovascular diseases. The combination of the electrocardiogram (ECG) symptoms, the echocardiographic images, and patient clinical history makes it possible to find out complex cardiac abnormalities at an early stage. In case of arrhythmias and heart failure, combining the time-series ECG data with the image-based structural data is beneficial in diagnosing the disease as presented in table 3 data. Moreover, EHR data may be used in remote monitoring and early intervention by combining wearable sensor data containing real-time heart rate and activity patterns with EHR data. These systems are especially useful in preventive healthcare, where hospitalization can be minimized and the associated expenses can be minimized by the timely identification of abnormalities. Multimodal analysis helps in neurological disorders such as the Alzheimer disease, Parkinson and epilepsy. Combination of neuroimaging (e.g., functional MRI), electrophysiological (EEG), and records of cognitive assessment makes it possible to diagnose the disease in a young age and monitor the progression of the disease. Multimodal models are able to identify fine-tuning changes in the brain structure and functioning which might not have been noticeable in one of the modalities.

VII. Performance Evaluation and Benchmarking

The evaluation of the effectiveness, reliability and clinical applicability of multimodal machine learning systems is very important as performance evaluation is a component. Since unifying heterogeneous biomedical data is highly complex, powerful evaluation systems will be needed to not only evaluate predictive quality but also the role of each modality, the ability of the model to generalize and the practical applicability of this model in the real world. The all-encompassing benchmarking strategy will make sure that the multimodal approaches are strictly checked on the basis of the set baselines, as well as clinical norms. The analysis of multimodal models is usually initiated by conventional classification and regression measures. Accurate, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC) are some of the commonly used metrics in diagnostic tasks.

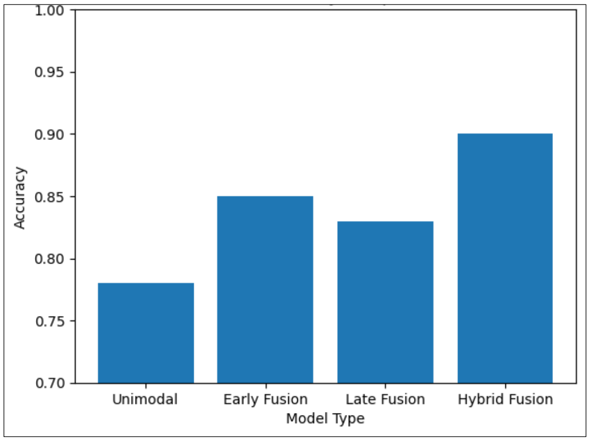

Figure 4. Comparison of Classification Accuracy Across Unimodal & Multimodal Fusion Strategies

Along with conventional measures, multimodal systems shall also require special measures in order to assess cross-modal interactions and robustness as shown in figure 4. One such aspect is the modality contribution analysis that analyses the contribution of each source in the model performance. Another method such as ablation studies is also common where one of the modalities is ablated to observe the change in performance. This aids in determining the most informative modalities and makes sure that the model is useful in capitalizing on complementary data. In addition, deep learning models can be trained with attention that is weighed to provide interpretability, or show how significant each modality is in decisions.

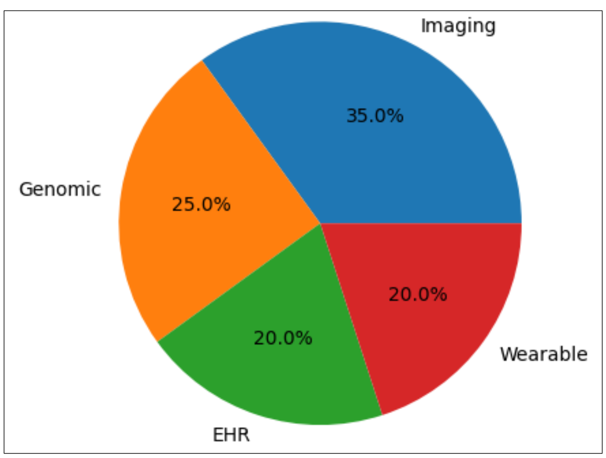

Figure 5. Relative Contribution of Different Biomedical Data Modalities to Model Performance

The other benchmarking technique is to compare multimodal models with multimodal model to the unimodal and traditional machine learning models as shown in figure 5. The comparisons of this sort illustrate the extra value of multimodal integration towards predictive strength and soundness. To achieve reproducibility and standardization, biomedical datasets, such as imaging repositories, genomic databases, and EHR-based cohorts are publicly available. Cross-validation methods (e.g., k-fold validation) are used to reduce overfitting, and make sure that models are generalizable to other patient groups.

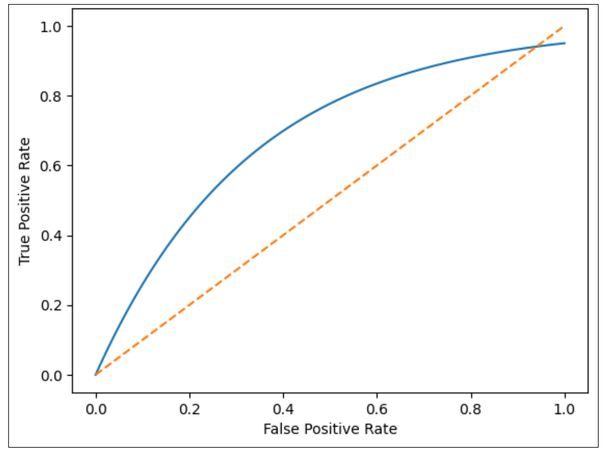

Figure 6. Receiver Operating Characteristic (ROC) curve illustrating the performance of the multimodal model.

Performance evaluation is a critical factor like scalability and computational efficiency, especially in the big biomedical applications as shown in figure 6. In practical health care settings, missing data is the norm and models need to be able to perform even when modalities are missing. The analysis of multimodal systems under simulated missing-data conditions is useful in examining the resilience of systems. Moreover, it is important to analyze fairness and bias to make sure that models are consistent in their performance with regard to various demographic groups and, therefore, contribute to fair healthcare provision.

VIII. Conclusion

Multimodal machine learning has become a tool that has the potential to unite heterogeneous biomedical data and develop intelligent healthcare systems. It can integrate various data, including medical imaging, genomic data, electronic health record and wearable sensor data to provide a complete picture of the health of the patient. This combination helps to improve the accuracy of the diagnosis, enables effective prognostic modelling, and helps to achieve personalized treatment plans based on precision medicine. Various fusion strategies have an important role to play in this process; early fusion in capturing feature-level interactions, late fusion in offering modularity and robustness to capture cross-modal relationships, and hybrid fusion balances the two to exploit cross-modal relationships effectively. An effective system architecture is critical towards a smooth flow of data, effective preprocessing and scalable learning. Multi-layered architecture assists interoperability, computational effectiveness and real-time implementation within the clinical setting. This evaluation performance shows that multimodal models are always more effective than unimodal models, especially when trained with the help of the state-of-the-art deep learning models like transformers and multimodal autoencoders. Multimodal learning is clinically significant as it has been used in oncology, cardiovascular medicine, neurology and critical care. The use of complementary modalities in the diagnostic process not only makes the diagnosis more accurate but also aids in the decision-making process in a timely manner. Although these benefits exist, there are challenges that exist which include: data heterogeneity, missing modalities, interpretability, privacy issues and system integration, which need to be overcome to be widely used in clinical environments. The areas of future development will be explainable artificial intelligence, federated learning to share data securely, and standard benchmarking systems. The capabilities will be further improved with integration with new technologies like digital twins and IoT-based healthcare systems. Ongoing researcher, clinician, and policymaker partnerships will be necessary to achieve patient-centric scalable healthcare solutions.

References

[1] Albahra, S., Gorbett, T., Robertson, S., D'Aleo, G., Kumar, S. V. S., Ockunzzi, S., Lallo, D., Hu, B., and Rashidi, H. H., "Artificial intelligence and machine learning overview in pathology and laboratory medicine: A general review of data preprocessing and basic supervised concepts," *Semin. Diagn. Pathol.*, vol. 40, pp. 71–87, 2023.

[2] Sansone, M., Fusco, R., Grassi, F., Gatta, G., Belfiore, M. P., Angelone, F., Ricciardi, C., Ponsiglione, A. M., Amato, F., Galdiero, R., *et al.*, "Machine learning approaches with textural features to calculate breast density on mammography," *Curr. Oncol.*, vol. 30, pp. 839–853, 2023.

[3] Das, S., Karanam, R. C., Krishna, O. B., and Maiti, J., "Machine learning and data analysis based breast cancer classification," in *Machine Learning Algorithms and Applications in Engineering*. Boca Raton, FL, USA: CRC Press, 2023, pp. 117–129.

[4] Saleh, G. A., Batouty, N. M., Haggag, S., Elnakib, A., Khalifa, F., Taher, F., Mohamed, M. A., Farag, R., Sandhu, H., Sewelam, A., *et al.*, "The role of medical image modalities and AI in the early detection, diagnosis and grading of retinal diseases: A survey," *Bioengineering*, vol. 9, p. 366, 2022.

[5] Popescu, D., Stanciulescu, A., Pomohaci, M. D., and Ichim, L., "Decision support system for liver lesion segmentation based on advanced convolutional neural network architectures," *Bioengineering*, vol. 9, p. 467, 2022.

[6] Najjar, R., "Redefining radiology: A review of artificial intelligence integration in medical imaging," *Diagnostics*, vol. 13, p. 2760, 2023.

[7] Shaik, T., Tao, X., Li, L., Xie, H., and Velásquez, J. D., "A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom," *Inf. Fusion*, vol. 102, p. 102040, 2024.

[8] Maleki Varnosfaderani, S., and Forouzanfar, M., "The role of AI in hospitals and clinics: Transforming healthcare in the 21st century," *Bioengineering*, vol. 11, p. 337, 2024.

[9] Wang, Y., Yin, C., and Zhang, P., "Multimodal risk prediction with physiological signals, medical images and clinical notes," *Heliyon*, vol. 10, p. e26772, 2024.

[10] Yuan, M. *et al.*, "Large language models illuminate a progressive pathway to artificial intelligent healthcare assistant," *Med. Plus*, vol. 1, p. 100030, 2024.

[11] T. S. Nipane, P. R. Varathi, D.M.Chaudhari , Y. V. Jivtode, and M. Naralkar, "Machine Learning for Predictive Analytics: Application in Healthcare", IJACECT, vol. 14, no. 1, pp. 328–330, May 2025.

[12] Behrad, F., and Saniee Abadeh, M., "An overview of deep learning methods for multimodal medical data mining," *Expert Syst. Appl.*, vol. 200, p. 117006, 2022.

[13] Meskó, B., "The impact of multimodal large language models on health care's future," *J. Med. Internet Res.*, vol. 25, 2023.

[14] van der Velden, B. H. M., Kuijf, H. J., Gilhuijs, K. G. A., and Viergever, M. A.,

"Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, p. 102470, 2022.
[15] Hulsen, T., "Literature analysis of artificial intelligence in biomedicine," *Ann. Transl. Med.*, vol. 10, p. 1284, 2022.
[16] Gu, J., Gao, C., and Wang, L., "The evolution of artificial intelligence in biomedicine: Bibliometric analysis," *JMIR AI*, vol. 2, 2023.
[17] Pei, X., Zuo, K., Li, Y., and Pang, Z., "A review of the application of multi-modal deep learning in medicine: Bibliometrics and future directions," *Int. J. Comput. Intell. Syst.*, vol. 16, p. 44, 2023.
[18] Barua, A., Ahmed, M. U., and Begum, S., "A systematic literature review on multimodal machine learning: Applications, challenges, gaps and future directions," *IEEE Access*, vol. 11, pp. 14804–14831, 2023.