Indian Journal of Computer Science and Technology
Special Issue, Volume 5, Issue 1 (January-April 2026), PP: 85-91.
https://www.doi.org/10.59256/indjcst.20260501C014
International Conference on Cyber and AI Security (ICCAIS-2026)
www.indjcst.com
ISSN No: 2583-5300

Transforming Healthcare Diagnostics through Multimodal AI

Jayaguru sahu[1], Ayush Kumar Behera[2], Soumya Ranjan Mishra[3], Sachikanta Dash[4], Pedina Bavishya[5], Suvam Nahak[6]
[1, 2,3,5,6] *Department of Computer Science Application, GIET University Gunupur, India.*
[4]*Madanapalle Institute of Technology and Science University, Andhra Pradesh, India.*

To Cite this Article: *Jayaguru sahu[1], Ayush Kumar Behera[2], Soumya Ranjan Mishra[3], Sachikanta Dash[4], Pedina Bavishya[5], Suvam Nahak[6] , "Transforming Healthcare Diagnostics through Multimodal AI, Indian Journal of Computer Science and Technology Volume 05, Issue 01 (January-April 2026), PP: 85-91.*

Abstract: *Artificial intelligence systems demonstrate their advanced capabilities in healthcare applications through their recent development of multimodal techniques which enable better diagnostic assessment and clinical decision-making processes. The system employs diverse statistics assets which encompass clinical imaging and electronic fitness statistics and physiological indicators and scientific documentation to create a complete model that correctly portrays the complex nature of scientific conditions. The field has developed several multimodal artificial intelligence systems for medical applications although research studies have not yet established a comprehensive framework that shows their complete operational capabilities. The research review presents an extensive assessment of all machine learning and deep learning multimodal systems which exist today while showing their different methods of data synthesis and their effectiveness in diagnostic assessments. The research study begins with a presentation of accessible datasets to the study team which will analyze the medical data preprocessing methods that establish uniformity across different types of medical information. The research study proceeds to define major fusion techniques which enable the combination of distinct information sources from various modalities and it presents typical model designs which extend from hybrid systems to transformer-based vision-language models and optimization-based educational resources systems. The current research studies show multiple operational problems which need to be solved. Our research shows that using multiple data sources for diagnosis results in better accuracy and system durability and capability to apply findings to new situations than using single data sources. The research presents an integrated framework which shows current research in the field while it creates new research paths to develop multimodal diagnostic system that medical professionals can use with practical understanding and operational efficiency.*

Key Words: *Multimodal AI, ML, DL, Fusion Techniques; medical diagnostics, DL.*

I. INTRODUCTION

The rising range of digitized medical records permits artificial intelligence (AI) [1] to decorate the technique of creating scientific selections. Medical diagnostics have historically trusted unimodal facts sources, inclusive of radiological images [2], clinical notes, or physiological alerts which physicians evaluated as separate entities. Unimodal systems can handle most medical cases successfully. Unimodal system cannot fully illustrate how human diseases develop because actual human diseases show all their symptoms through different types of medical evidence.

Accurate diagnosis and treatment of cancer, dementia, heart disease, and metabolic disorders require physicians to consider many types of medical evidence. Accurate prognosis and treatment of most cancers, dementia, coronary heart ailment, and metabolic problems require physicians to do not forget many sorts of scientific evidence. Multimodal Artificial Intelligence (MAI) operates as a technique to those demanding situations because it allows organizations to mix distinct styles of facts which incorporates photographs and signals and dependent information and medical narratives right into an unmarried machine for evaluation. The structures use the unique talents of every modality to beautify their diagnostic capabilities and their capacity to are expecting effects and their potential to interpret medical effects.

The conducted studies that provided particular implications regarding fusion taxonomy and medical QA structures and transformation of large language models into fairness and bias assessment and multimodal frameworks. The studies demonstrates3 developments which consist of standardized evaluation methods and general-use architectural designs and ongoing problems that involve machine growth and knowledge and fair system distribution. The observe by way of indicates how multimodal synthetic intelligence technology benefits oncology studies through its packages in modellingtumor heterogeneity and remedy reaction evaluation for B-mobile non-Hodgkin lymphoma patients. MAI generation indicates super overall performance upgrades in medical fields that consist of tumor type dementia subtyping fetal threat assessment and important care prognosis. Medical testing now includes multimodal artificial intelligence as it works with both transportable diagnostic equipment and coffee-energy sensor fusion systems and adaptable affected person tracking structures. The research investigates how thermal sensors and ultrasound sensors and SAR sensors can work together with lightweight neural networks for diagnostic system development [11]. Researchers

demonstrate how hybrid soft- computing methods can be used for both intermediate fusion and benchmarking [12], while they describe how portable patient monitoring systems need adaptive feedback mechanisms for their operation [13].

The MAI system now expands its operational capabilities through its new mobile functions which deliver quick and adaptable clinical assessment tools. It provides comprehensive explanations of the basic issues which affect multimodal learning. The primary obstacles which need to besolved include six different areas which are representation and alignment and inference and generation and transference and quantification. Unbalanced data and specific noise types which affect different modalities together with the need for explanation create extra difficulties for the clinical field. The evaluate demonstrates that studies on MAI needs standardized fusion frameworks and area-precise benchmarks which presently do not exist, hence blockading its pathway to medical implementation. It carried out a cutting-edge survey which indicates how medical responsibilities want higher integration, while they advise that destiny frameworks need to improve MAI systems thru higher temporal dependency, personalization, and gadget performance monitoring. Deep getting to know primarily based multimodal artificial intelligence systems [16] have won reputation due to the fact they are able to extract hierarchical features from both unstructured and established records. The research shows that models which combine transformers and convolution networks with hybrid attention mechanisms perform effectively in tasks related to disease classification, segmentation and retrieval and risk prediction [3,17].

The existing challenges require solutions because they involve missing modalities and institutional generalization and regulatory approval processes. The research review examines current developments in multimodal artificial intelligence technology which is used for medical diagnostics. The research paper presents a multimodal dataset taxonomy, analyses preprocessing methods which enable modality integration, describes different fusion techniques and evaluates the performance of various model architectures. The study shows how technological progress has created new opportunities for clinical use of MAI while also presenting challenges to its interpretability and demonstrating its actual effects on medical practice.

This review studies the latest developments in multimodal artificial intelligence which researchers use to create medical diagnostic systems. The study provides a multimodal dataset classification system which it uses to assess various pre-processing methods that unify different data types and to evaluate different data integration techniques and their impact on model performance and architectural design. The study demonstrates how MAI technology has developed throughout time while showing the difficulties of understanding its use in real clinical situations.

Multimodal artificial intelligence functions as a contemporary healthcare technology that introduces an innovative approach to medical treatment. The conventional machine studying technique relies upon on separate records resources which create an incomplete photograph of an affected person's fitness condition. The system of fragmented statistics impacts bad results, which are not relevant to the treatment of scientific matter on time and misleading diagnoses and unreliable forecasts of outcomes. In conjunction, incorporated digital health record designs [18], handheld devices, radiology information and biomedical designs give a rich foundation in the creation of MAI. However, the ability is not exploited to the full amount because of the fact that there are no standardized tools that can paintings that have one of a kind assets and method complex data. The research proposes the extended want to have device structures that can be capable of incorporatingmultiple facts assets appropriately and still maintain the interpretability and the consistent overall performance of the medical care. Multimodal artificial intelligence creates clinical advantages because it enhances diagnostic accuracy and enables customized patient treatment and streamlines labor-intensive tasks. The medical field has adopted integrated learning methods because doctors need to combine multiple diagnostic tools which include imaging and laboratory tests and patient medical history. The gap between clinical reasoning and algorithmic inference gets reduced through MAI implementation.

This evaluate gives a dependent overview which mixes findings from a couple of latest research about multimodal artificial intelligence development in medical diagnostic systems. The literature review presents multiple research perspectives through which we analyses current studies about multimodal datasets which have been used in recent scientific investigations. The study examines preprocessing methods which help enhance data quality and enable different modalities to work together through three main techniques: normalization and resampling and feature selection. The study presents several multimodal fusion methods that use initial fusion and mid-level feature linking and cross-modal attentional mechanisms to create matching representations. The take a look at offers its content through separate sections which demonstrate the multimodal studying process that begins with records collection and fusion and ends with version development and implementation. Our study provides a summary of critical findings while we present existing obstacles and potential research paths which scientists can explore within this fast-developing research domain. The review compares different methods used in various studies while showing how model performance has progressed and stressing the need for methods that can be used in clinical settings and demonstrate performance in various situations.

II.METHODOLOGY

The researchers used titles to check if papers matched their requirement for multimodal learning and diagnostic research. The researchers needed to check both the abstracts and the complete texts to establish whether the study employed several data types and their results were applicable to medical diagnosis and prediction. The researchers eliminated articles from their study which either belonged to non-medical fields or studied only one type of data or contained content in languages other than English. The team deleted duplicate documents while keeping only those papers which had the greatest value for their research work. The tables and synthesis include only those studies which the present review directly cites and analyses. The study required original research that had received peer review and used two distinct data types and its results needed to show quantitative assessment of clinical diagnostic model performance. The researchers eliminated all case reports and editorials and research papers which did not use hybrid methods or which used only a single data type. The researchers established a comprehensive framework to select research articles which would show how multiple artificial intelligence systems work together with medical diagnosis. The research

team followed the PRISMA guidelines for their review process. The PRISMA go along with the glide diagram [19] in Fig. 1. explain stairs discovered for the duration of the identity thatintroduced about final choice of research considered on these paintings.

A.Multimodal Datasets

The development of multimodal datasets as research tools for medical diagnostics research requires datasets that combine diverse data types including images and clinical records and textual annotations. The datasets allow machine learning model development through their ability to process genuine healthcare data. The records were collected in the PAD-UFES-20 data set in 2018-2019 in its Dermatological and Surgical Assistance Program in Brazil and consists of 2298 medical pix with the associated metadata describing 1373 patients and 1641 pores and skin lesions (Pacheco et al., 2018). The documentation involves pix that enquires about the variety of desire options and one unmarried variety of approximately smartphones. The machine can include 6 diagnostic groups that may include basal cell carcinoma and squamous cellular carcinoma and actinic Keratosis and seborrheic Keratosis and Bowens disorder and majority of cancers and nevus-Scientific metadata will consist of 21 attributes that include age and intercourse and lesion area and pores and skin kind and lesion duration. Biopsy affirmation was done in about 58.4 percent of the lesions. The researchers ensured that they verified the picture resolution and complete facts requirements owing to the fact that they had to verify all scientific facts which they translated into English language.

The MedICaT dataset is better than the one by Subramanian and his research organization because of the fact that they gathered 217060 scientific photographs by searching 131410 biomedical studies articles. The data set contains two types of textual content statistics that consists of discern captions as well as inline references picked out of full-text articles, sub-determine, and sub-caption remarks. The dataset entails 75 percent composite numbers and 74 percent photographs which indicate in-line references enabling the researcher to conduct an extensive study on the manner in which snap shots interrelate with their text. The data allow scholars to art work on sub-determine-to-sub- caption matching responsibilities and parent retrieval. The system is viewed as a challenging evaluation device that determines the comprehension of seen medical question answering skills together with file stage image retrieval systems and medical article.

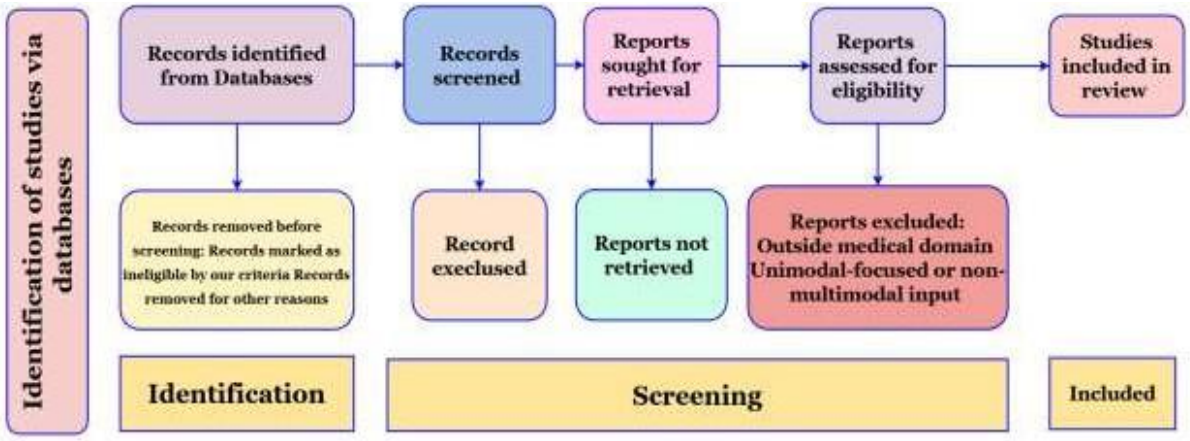

Fig. 1. Systematic Selection Process

The FFA-IR dataset serves as a number one benchmark which permits medical specialists to create reliable scientific evaluations through its descriptive content material. The dataset consists of 1330 fundus photographs that are matched with each Chinese and English diagnostic evaluations and extremely good lesion mapping. The sample consists of structured eye sickness tags which researchers manually aligned into textual descriptions and created bounding container annotations for commonplace lesions which includes micro aneurysms and hemorrhages. The ophthalmologists selected those photos and reviews due to the reality they desired to enhance research on explainable visible query answering and lesion-based totally grounding. The dataset allows record detail alignment with unique picture regions which allows multimodal models to undergo schooling and trying out through every generative and retrieval-based completely techniques.

The system supports multiple X-ray and CT and MRI modalities which enable users to examine various body parts and medical conditions. The dataset ROCOv2 extends the existing dataset through improved annotation methods which include an extra 11000 images from PMC-OAI content to provide more extensive data coverage. The package containsmultiple samples which contain thorough metadata details that include article headings and image captions together with medical terms.

The datasets determine standard performance requirements that researchers apply in order to assess go- modal retrieval and imaginative and prescient-linguistic anchoring and obligation in the development of clinical documents. Pediatric Radiology Dataset This dataset involves a hundred and eighty radiographs of pediatric cases in conjunction with clinically oriented pair-want diagnosis queries (MCQs). The dataset includes sample of different parts of the anatomy along with the chest, stomach, musculoskeletal machine and head. The machine supports research efforts that seek to investigate the effect of instructions and medical reasoning that is completely dependent on the pediatric imaginative and precognitive language commitments.

The SLAKE-VQA dataset is a bilingual clinical visual query-answering useful resource, which incorporates semantic information in English and Chinese. It includes 642 radiographs of 3 public repositories comprising of CT, MRI and X-ray modalities and an anatomical insurance of themind, neck, chest, stomach cavity and pelvic cavity. The information gives 14,028

pairs of questions and answers, each of the open-ended and closed-ended kinds, to evaluate clinical and visual reasoning issues, which include figuring out the modality, figuring out the organ, and detecting the abnormality. The QA shape is primarily based on a established medical information graph with more than 5,200 curated triplets via SLAKE. The dataset has been partitioned into three agencies, specifically, 9,849 schooling samples, 2,109 validation samples and 2,070 check samples. The National Alzheimer's Coordinating Center (NACC) records set [34] is a central location of standardized multimodal statistics that has been received in over 40 Alzheimer ailment research centers within the United States. It includes documents of over 19,000 sufferers that had been exposed to structural brain MRI scans, neuropsychological tests, bio specimen series and lengthy-time period clinical comply with-ups.

B. Data Preprocessing Techniques

Multimodal facts for gadget mastering models needs powerful preprocessing which serves as the essential primary requirement. Preprocessing multimodal medical information requires the method of cleaning and normalization together with the synchronization of a couple of capabilities which include photograph and timeseries and textual information [35]. Medical data requires preprocessing techniques which depend on the structural elements and resolution capabilities and semantic information of each source. The section presents an analysis of different studies which tackle the difficulties involved in processing multimodal medical data. The segment outlines the preprocessing workflows which deal with imaging and sign and textual content and tabular information and demonstrates their alignment and transformation and integration techniques before version training. The identified studies are organized into organizations that constitute the primary modality or task awareness that each study uses in its preprocessing tasks. The information obtained stratification based solely on clinical categories led to harmonizing processes that reduced variability among sites at the equipped appropriate management with the methods of file exclusion and variable sparse elimination and it was viewed by applying evaluation standardization and area heuristic value modeling and dimensionality rest. Thru z-rating normalization, [36] convoluted ADNI structural MR and PET snap shots through spatial resolution normalization. CSF biochemistry

Created a complete gadget which processes clinical data through CDW-H thru first securing affected person records via anonymization after which processing essential laboratory statistics in line with PCORnet model standards. The VGG-based totally definitely classifier decided on PLAX and A4C echocardiogram views at the same time as CNN layers produced embed dings which hobby pooling used. Eliminated CTG indicators from CTU-UHB which contained an awful lot less than 10,000 samples at the identical time as they used sparse dictionary studying to denies signs and GAN-primarily based augmentation to achieve class stability [37, 38].

The various multimodal study methods develop fusion methods which determine the process of combining different sources of information into a model. The Medical programmer in great it model which exists in this system, uses three different merging methods. To show how they extract functions and combine data and make predictions at excessive performance stages. The scientists from the research developed hybrid fusion models which combine multiple fusion techniques to predict results based on one modality while using another. The section provides a classification system that categorizes fusion methods based on their growing popularity, while it shows all research examples for each category.

III.MULTIMODAL APPROACHES AND MODEL ARCHITECTURES

Medical field has also utilized several system design versions to handle diverse medical data which includes multiple input types and various medical objectives. The section will provide a full study of the development of the current models that will be organized as per architectural design trends and modelling approaches that the Table three determines.

A. Hybrid and Attention-Based Architectures

Evolved fusion techniques through five great version designs which they used to discover cardiac amyloidosis. The researchers assessed 5 experimental structures thru their utility of precise techniques which resulted in growing various testing environments. The first approach required researchers to create a baseline machine that used first-class digital fitness records as input records. The 2nd technique worried comparing more systems that depended on distinct records resources. The have a look at assessed two structures that used visual information from PLAX and A4C for evaluation via a visible records fusion technique. Two systems used photograph embedding collectively with EHR information for seen information fusion. The baseline machine executed an AUROC score of 94.1% through its mixture of critical fusion techniques and flexible modality alignment techniques. The have a look at affords a visual precis which demonstrates five architectural designs collectively with their distinctive fusion strategies. Fig. 2. evaluates five architectural techniques for multimodal cardiac amyloidosis kind. The system desires three critical parts to carry out obligations which include radionics feature extraction from abdominal X-ray pix and feature selection through mRMR and LightGBM class that uses both radiographic information and medical statistics.

The selection technique for the version concluded with SENet-154 as the very last preference because it established amazing overall performance all through the assessment procedure at the same time as its function embedding served future prediction requirements. The effects which finished an AUC of 90 three factor 37% and an accuracy of 92.5%. The surgical prediction machine achieved a ninety-four. Thirteen% AUC value mixed with an 88.61% accuracy rate thru using each radiographic and scientific records which passed the results obtained from single-modality structures. The TDN model delivered an progressive massive-kernel extension which combines a pass-attention-primarily based fusion module. The BKTDN model become created to study eye-movement films due to the fact it may detect distinct eye-movement styles via its brief-term and lengthy-time period convolutional streams which use good sized receptive subject sizes. The system integrated head-motion information through its dual framework system which blended a self-encoder with a pertained model.

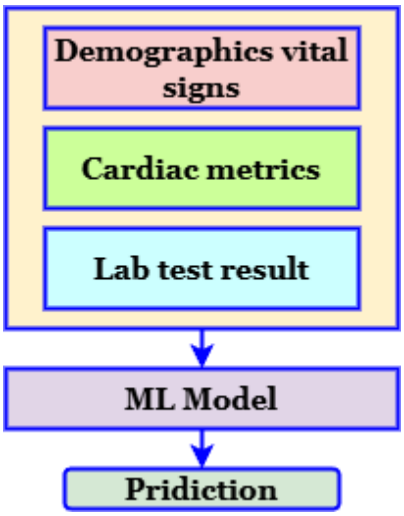

Fig. 2. Overview of model architectures

B. EHR-Centric and Optimization-Based Models

The multimodal diagnostic framework is shown in developed AutoFM, which features as a neural architecture seek platform that detects dependable fusion designs which integrate which makes use of EHR records and area knowledge with interest and calibration mechanisms to interpret records. The system uses a main controller component to build different architectural designs which the system tests through reinforcement learning to measure their full operational capacity. The machine uses AutoFM to choose which components from EHR domain names need to be mixed based totally on prognosis and laboratory testing, and scientific processes. The device operates as a longtime trendy answer which tested a hit effects for the duration of three separate sickness prediction checks that included coronary heart failure, diabetes, and mortality assessment[39]. The advanced architectural designs proven a couple of connectivity options with specific architectural intensity and fusion capabilities, which confirmed that scientific prediction obligations advantage from adaptive architecture selection. The version carried out AUROC consequences of 84.8 % for coronary heart failure and 85.7 percent for diabetes and ninety-one. Four percentage for mortality. Created a hybrid model which combines the ILHHO rule set with the KELM classifier gadget to beautify the accuracy of Alzheimer ailment predictions. The feedforward neural network processed all enter types through their wonderful encoding methods which created a single output that integrated all statistics into one visible illustration.

C.Tabular-Image Fusion Architectures

Advanced a cooperative framework which permits human beings to paintings collectively with artificial intelligence via their superior multimodal deep learning system, which achieves better diagnostic outcomes for distinct lupus erythematosus subtypes The gadget evaluated twin-scale multi-IHC snap shots together with medical pix and affected person medical facts via its twin-route evaluation system. The first course used ResNet-50 to evaluate IHC patches, on the identical time as the second path used EfficientNet-B1 to investigate clinical pics. The gadget used concatenation to mix exceptional modality embedding before the shared category head processed them. The multimodal version carried out an AUROC of eighty-4percentage which exceeded each the image-only and EHR- simplest models that finished AUROCs of 80% and 78% even as demonstrating the blessings of using more than one records types for making picks approximately excessive- hazard orthopedic sufferers.

D.General-Purpose LLMs and Instruction-Tuned Models

The research team led by means of Panagoulias evaluated the GPT-4V system which generated medical checks from both radiology and dermatology photos alongside clinical case descriptions. The machine established a new standard for image comprehension by achieving elevated RadGraph F1 scores and using Rouge-L to measure text similarity between the extracted elements and actual report content. The system showed its ability to operate on multiple datasets by achieving a RadGraph F1 score of 79.5% and a cosine similarity of 93.4% for pediatric radiographs, which proved that GPT-4V can support multimodal diagnostic systems, despite lacking specific clinical training materials. The basic overall performance traits of zero-shot systems, in comparison with pleasant-tuned models pattern [40].

The researchers developed a self-supervised multimodal learning system which they built on the MIMIC-CXR dataset that included vast amounts of raw medical image and report data. The system achieved effective X-ray image and report extraction through its ACNN-based visual encoder and transformer-based text encoder which used contrastive methods to obtain matching of these two types of data. The system shared a common representation space which enabled multiple tasks to be executed after the pre-training phase through zero-shot radiographer classification and report understanding tasks. The model executed an average AUROC score of 78.1% at some stage in 0-shot testing at the ChestX-Ray14 benchmark. The researchers assessed ChatGPT-4V which functions as a vision-to-speech version that became pre-skilled on hooked up pathology duties in keeping with Zhu et al. The gadget operates as an LMM black area which methods visual-language facts via its widespread transformer framework. The device recognized multiple elegance of visual content material thru its multiple content material identification skills.

IV.RESULT & DISCUSSION

The existing multimodal AI systems demonstrate their diagnostic capabilities in different scientific areas which include cardiology and neurology and oncology and maternal-fetal health. The success of these systems depends on the selection of appropriate fusion techniques together with the implementation of specific preprocessing methods for different modalities. The researchers used 3 distinct strategies information produced better consequences than simpler fusion strategies. The researchers used SENet- primarily based radiomic functions together with clinical records to expand a version which predicted NEC surgical eligibility and accomplished an AUC rating of 95.3%. The aggregate of intermediate and selection-degree fusion methods together with domain-precise preprocessing carried out progressed diagnostic outcomes. The business operations of neurological applications produce monetary returns that match their operational costs. The researchers used system which enabled them to gain 99.2% achievement in identifying Alzheimer's sickness.

The implementation of multimodal diagnostics encounters difficulties because the required information remains either unavailable or insufficient. Researchers use overdue fusion or selection-level ensembling to solve problems caused by incomplete data. maintain their operational capabilities through adaptable modality alignment and contrasting pretraining methods. The BKTDN model developed by Lu and his team lost approximately 30 percent of its accuracy when researchers removed head vectors which demonstrated that performance depends on specific modalities. The system accuracy decreases when information sources provide conflicting or unreliable data which demonstrates the necessity for developing robust processing methods. The benchmarking matrices and interpretability tools hold essential importance. GradCAM and saliency maps analyze imaging data while SHAP and feature attribution methods evaluate EHR and tabular records. Used saliency and attribution analysis for dementia analysis, Combined IHC and photographic heatmaps to aid human information. The medical field can use the outputs from these two systems.

The current MAI systems deliver effective results yet their complete interpretability testing needs further development. The models show excellent performance but their failure assessment and risk evaluation across different domains remain restricted. The main obstacles contain organizational data overfitting together with environmental disturbances and the insufficient evaluation process across multiple affected patient groups. The research needs further development to resolve the conflict which exists between different data sources according to their needed patient safety and complete assessment of their diagnostic methods. The general-purpose LLMs which include ChatGPT-4V create new reproducibility problems together with hallucination issues and domain-specific hazards which affect critical fields such as histopathology and surgical trial procedures. The adaptive MAI systems progress through clinician feedback and patient-specific information which they collect over time. The systems enable prediction modifications based on contextual information together with weight updates. The researchers developed an AI system which uses multiple data sources to monitor patients continuously through sensors and medical information, enabling medical professionals to provide customized treatment. The implementation of effective patient group processing and fusion methods through multimodal AI leads to superior diagnostic results. Standardized benchmarks are required to set testing requirements which will guarantee model transparency and enable academic institutions to measure their models through standard performance evaluation. Future research needs to establish solutions which will handle the ethical and regulatory and operational challenges that arise from patient consent and responsibility and model management. The new methods of guidance tuning and contrastive pertaining and unsupervised learning show potential to develop clinical AI systems which will operate effectively and provide transparent and trustworthy results.

V.CONCLUSION

The current multimodal artificial intelligence research studies in clinical diagnostic applications which demonstrated that integrating imaging data with electronic health records and crucial clinical information and unstructured text medical documents results in improved disease detection and better clinical decision support. The MAI systems demonstrate superior performance compared to single-modality systems throughout various medical fields which include cardiology and neurology and oncology and radiology and pediatrics. The studies confirmed that corporations now use hybrid strategies which combine separate intermediate techniques with hobby-primarily based techniques for their scientific imaging paintings that uses self-supervised vision-language models based totally on BLIP-2 and GPT-4V era. The studies established a collection of public and institutional datasets which blanketed precise preprocessing strategies that researchers used to deal with each dataset kind. The research achieved successful results when EHRs were combined with imaging systems because predicted models successfully predicted essential medical outcomes which included sepsis and surgical risk and death. The research identified multiple obstacles which included challenges with data standardization and difficulties with machine learning model evaluation and problems with designing user interfaces.

REFERENCES

1. Simon, B.; Ozyoruk, K.; Gelikman, D.; Harmon, S.; Türkbey, B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: A narrative review. Diagn. Interv. Radiol. 2024.
2. Laganà, F.; Bibbò, L.; Calcagno, S.; De Carlo, D.; Pullano, S.A.; Pratticò, D.; Angiulli, G. Smart Electronic Device-Based Monitoring of SAR and Temperature Variations in Indoor Human Tissue Interaction. Appl. Sci. 2025, 15, 2439.
3. Mishra, S. R., Dash, S., Padhy, S., Kumar, N., & Dash, Y. (2024, September). Integrating Multi-Omics Data for Advanced Diabetes Prediction and Understanding. In 2024 7th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 7, pp. 1447-1453). IEEE.
4. Dash, S., Mishra, S. R., & Baboo, A. (2025, January). Enhancing Diabetes Prediction using Hybrid Ensemble Approach. In 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) (pp. 1-6). IEEE.
5. Mario, V.; Laganá, F.; Manin, L.; Angiulli, G. Soft computing and eddy currents to estimate and classify delaminations in biomedical device

CFRP plates. J. Electr. Eng. 2025, 76, 72–79.
6. Menniti, M.; Laganà, F.; Oliva, G.; Bianco, M.; Fiorillo, A.S.; Pullano, S.A. Development of Non-Invasive Ventilator for Homecare and Patient Monitoring System. Electronics 2024, 13, 790.
7. Dash, S., Mishra, S. R., & Baboo, A. (2025, January). Efficient Prediction of Diabetes Mellitus Through Hybrid Ensemble Machine Learning Model Using IoT. In 2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET) (pp. 1- 6). IEEE.
8. Mishra, S. R., & Dash, S. (2024, December). Machine Learning Based Diabetes Prediction Using the PIMA Indian Dataset. In 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1-6). IEEE.
9. Dash, S., Padhy, S., Kumar, N., & Nayyar, A. (2026). Medisecure: a hybrid approach for enhancing multimedia data protection in healthcare. Cluster Computing, 29(1), 75.
10. Li, Y.; Daho, M.E.H.; Conze, P.H.; Zeghlache, R.; Le Boité, H.; Tadayoni, R.; Cochener, B.; Lamard, M.; Quellec, G. A review of deep learning-based information fusion techniques for multimodal medical image classification. Comput. Biol. Med. 2024, 177, 108635.
11. Mishra, S. R., Dash, S., Padhy, S., & Das, R. K. (2024, September). Diabetic Foot Ulcer Diagnosis Through Deep Learning Model. In 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit) (pp. 1194-1199). IEEE.
12. Mishra, S. R., & Dash, S. (2024). Iterative Model Design for Diabetes Analysis Using FedOmics Causal Network and Federated Multi-Omics Variational Autoencoder. International Journal of Emerging Technologies and Innovative Research, 11(6).
13. Ruckert, J.; Bloch, L.; Brungel, R.; Idrissi-Yaghir, A.; Schäfer, H.; Schmidt, C.S.; Koitka, S.; Pelka, O.; Ben, A.; Abacha, A.G.; et al. ROCOv2: Radiology Objects in COntext Version 2, an Updated Multimodal Image Dataset. Sci. Data 2024, 11, 688.
14. Li, Q.; Yang, Z.; Chen, K.; Zhao, M.; Long, H.; Deng, Y.; Hu, H.; Jia, C.; Wu, M.; Zhao, Z.; et al. Human-multimodal deep learning collaboration in 'precise'diagnosis of lupus erythematosus subtypes and similar skin diseases. J. Eur. Acad. Dermatol. Venereol. 2024, 38, 2268–2279.
15. Mishra, S. R., & Dash, S. (2024). Predictive Analysis On Diabetes Detection Using Pima Indian Diabetes Dataset. IJRAR-International Journal of Research and Analytical Reviews (IJRAR), 11(2).
16. Baboo, A., Mishra, S. R., & Dash, S. (2024, November). An Improvised Diabetes Prediction System Using Hybrid Ensemble Approach. In 2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-6). IEEE.
17. Reith, T.P.; D'Alessandro, D.M.; D'Alessandro, M.P. Capability of multimodal large language models to interpret pediatric radiological images. Pediatr. Radiol. 2024, 54, 1729–1737.
18. Martin, S.A.; Zhao, A.; Qu, J.; Imms, P.E.; Irimia, A.; Barkhof, F.; Cole, J.H.; Initiative, A.D.N. Explainable artificial intelligence for neuroimaging-based dementia diagnosis and prognosis. medRxiv 2025.
19. Mishra, S. R., & Dash, S. (2026). AI - Driven Remote Health Monitoring for Predicting Diabetes and Heart Diseases Using ULMCSO and PGND Models. Hyper - Intelligent Networks: Exploring the Future of Connectivity for Society 5.0, 219-247.
20. Mishra, S. R., Dash, S., Padhy, S., & Samuel, P. (2026). Legal Aspects of Operating IoMT Applications in the Fog Computing. In Integrating Cloud, Fog, and Edge Computing in Healthcare: Federated Learning and Blockchain Approaches: Harnessing Distributed Technologies for Enhanced Healthcare Delivery (pp. 211- 224). Cham: Springer Nature Switzerland.
21. Xue, C.; Kowshik, S.S.; Lteif, D.; Puducheri, S.; Jasodanand, V.H.;
22. Zhou, O.T.; Walia, A.S.; Guney, O.B.; Zhang, J.D.; Pham, S.T.; et al. AI-based differential diagnosis of dementia etiologies on multimodal data. Nat. Med. 2024, 30, 2977–2989.
23. Schilcher, J.; Nilsson, A.; Andlid, O.; Eklund, A. Fusion of electronic health records and radiographic images for a multimodal deep learning prediction model of atypical femur fractures. Comput. Biol. Med. 2024, 168, 107704.
24. Rath, L., Mishra, S. R., Dash, S., Pradhan, P. C., & Baboo, A. (2025). Predicting diabetic patients coronary artery calcium score, deep learning using retinal images. In Intelligent Computing Techniques and Applications (pp. 113-118). CRC Press.
25. Pattnayak, A. P., Mishra, S. R., Dash, S., & Baboo, A. (2025). Utilization of deep learning and machine learning models to approach high glucose and low glucose prediction with type 1 diabetes mellitus in adult patients. In Intelligent Computing Techniques and Applications (pp. 102-107). CRC Press.
26. Dash, A. B., Dash, S., Padhy, S., Kumar, N., Pati, G. K., & Uthansingh, K. (2025). Leveraging inception-v3 CNN model for efficient image classification. In Intelligent Computing and Communication Techniques (pp. 341-348). CRC Press.
27. Niu, S.; Ma, J.; Bai, L.; Wang, Z.; Guo, L.; Yang, X. EHR-KnowGen: Knowledge-enhanced multimodal learning for disease diagnosis generation. Inf. Fusion 2024, 102, 102069.
28. Dora, N., Dash, S., Baboo, A., & Mishra, S. R. (2025, August). Efficient Nail Disease Diagnosis Using Deep Neural Networks for Predicting Abnormalities. In 2025 International Conference on Next Generation of Green Information and Emerging Technologies (GIET) (pp. 1-5). IEEE.
29. Mishra, S. R., Dash, S., & Rath, L. (2024, November). Effective Diabetes Mellitus Prediction Using a Hybrid Ensemble Machine Learning Model with Iot. In 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS) (pp. 1- 8). IEEE.
30. Dash, A. B., Dash, S., Padhy, S., Mishra, B., & Paikaray, B. K. (2025). Streamlining colorectal cancer diagnosis: leveraging MobileNet-V3 for efficient image classification. Int. J. Internet Manufacturing and Services, 11(4), 317 Zeng, L.; Ma, P.; Li, Z.; Liang, S.; Wu, C.; Hong, C.; Li, Y.; Cui, H.;
31. Li, R.; Wang, J.; et al. Multimodal Machine Learning-Based Marker Enables Early Detection and Prognosis Prediction for Hyperuricemia. Adv. Sci. 2024, 11, 2404047.