International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 1055- 1059
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0151

Explainable Multimodal Deep Learning in Healthcare: A Survey of Current Approaches

B. Dhanalaxmi[1], Dr. PVVS Srinivas[2]
[1]Research Scholar, Department of Computer Science &Engineering, KL University Vijayawada, India.
[2]Professor, Department of Computer Science &Engineering, KL University Vijayawada, India.
Emails: *Shikari22@gmail.com[1], cnu.pvvs@kluniversity.in[2]*

Abstract

Multimodal data integration has been considered the next step in transformation for modern healthcare as it brings an improved level of patient outcome and clinical decision-making. With the multimodal data set consisting of medical images, electronic health records, wearable sensor data, genetic information, and behavioral insights, the complexity of patient health becomes much clearer. Traditional methods for data analysis find it challenging in handling such complexities and diversities in data sets. This paper proposes a deep learning multimodal framework that exploits feature extraction, optimal selection of feature, and explainable AI techniques in order to detect and predict diseases. Data fusion techniques are used in the suggested system to efficiently combine various data sources, improving diagnosis accuracy and dependability. Furthermore, by using explainable AI techniques, the model guarantees decision-making transparency and helps doctors comprehend the roles that various modalities play in diagnostic results. Using a Python implementation on this framework brings promising results of disease categorisation and prediction with the possibility for AI-driven multimodal healthcare to improve medical diagnosis and individual therapy.

Keywords: *Multimodal Learning; Healthcare AI; Medical Data Fusion, Deep Learning, Electronic Health Records (EHR), Disease Prediction, Explainable AI (XAI), Medical Imaging, Sensor Data Analysis and Data-driven healthcare.*

1. Introduction

Artificial intelligence and machine learning have helped to revolutionize healthcare by bringing to the light of day unimaginable volumes of data and unprecedented complexity.[2] Traditionally, medical diagnosis rested on unimodal data sources, such as laboratory reports, or imaging modalities like MRI and CT scan. However, with the expanding multimodal nature of data types ranging from well-structured Electronic Health Records, to unstructured sensor-based or genomic data necessitates the advance development of computer models that should efficiently extract insightful knowledge [1]. Multimodal data are prone to complexity due to diversity, integration and computational overhead problems [1]. The inability to standardize multimodal fusion with missing data and class imbalance issues has limited the performance of traditional methods [1]. Recent breakthroughs in deep learning include robust feature extraction and classification techniques as well as valuable fusion techniques, thus providing more accurate and holistic models for diagnostics [2]. This paper presents a deep learning-based framework for combining multimodal health data to improve the detection and prediction of diseases[1][2].Using sophisticated feature selection, the model will extract the most important attributes from multiple modalities, enhancing the predictive power of the healthcare applications[2].Use of Explainable AI will guarantee the transparency and interpretability of model decisions, fulfilling the core necessity for trust in AI-driven healthcare systems[2].The proposed system bridges the gap between multimodal data integration and practical clinical applications, ultimately improving patient

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 1055- 1059
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0151

care and medical decision-making[1][2].

2. Literature Review

The healthcare context has increasingly embraced the integration of multimodal data as a revolutionary approach to improving patient outcomes and decision-making. Some studies discuss different data modalities, feature extraction techniques, and classification models in medical diagnosis and prediction. Researchers who authored the report "WatMIF: Multimodal Medical Image Fusion-Based Watermarking for Telehealth Applications" addressed a secure multimodal medical image fusion-based algorithm for watermarking. Kedar Nath Singh et al. elaborated on incorporating multiple medical image modalities that include MRI along with CT scanning for improved clinical diagnosis [3]. The study makes it clear as to how data fusion techniques enhance data security and robustness in telehealth applications, showing the importance of multimodal integration in medical data analysis. In a similar manner, Grzegorz Jacenków and others in their paper "Indication as Prior Knowledge for Multimodal Disease Classification in Chest Radiographs with Transformers" investigate the application of multimodal learning in disease classification [4]. Their experiment used textual indications and chest radiographs with transformer-based architectures to classify. It is demonstrated how combining the context of the text with imaging data improves diagnostic performance compared to unimodal models. In "Deep Neural Ensemble Classification for COVID-19 Dataset, " Fauzan Iliya Khalid et al. proposes a new approach of deep learning, integrating various forms of classifiers to achieve higher accuracy in COVID-19 detection [5]. The authors discuss the application of ensemble learning as a suitable strategy for task multimodal classification. The results show that the implementation of a combination of classifiers, including SVM, NB, and decision trees, exhibits higher prediction accuracy while processing multimodal medical datasets. Gan Cai and colleagues explore deep learning models for skin disease classification in "A Multimodal Transformer to Fuse Images and Metadata for Skin Disease Classification" [6]. Their work presents a neural network architecture that integrates clinical metadata with imaging data to improve diagnostic performance. The study supports the notion that multimodal approaches lead to superior classification outcomes in medical applications. Mayur Mallya and Ghassan Hamarneh have discussed the application of XAI in the field of healthcare in "Deep Multimodal Guidance for Medical Image Classification" [7]. This study explored the role of XAI methods in ensuring the transparency and trust in AI-driven medical diagnostics. With the multimodal datasets like histopathology images and MRI scans, this study shows that the explain ability further enhances the interpretability of the AI model for better clinical decision-making. Zhang et al., in the "Tomato Disease Classification and Identification Method Based on Multimodal Fusion Deep Learning," have detailed how multimodal fusion learning methods can be useful in the case of agricultural diseases diagnosis [8]. While more focused on classifying plant disease, this gives a valuable introduction to methodologies learned in multimodal learning, with potential for adaption into applications in healthcare. The study by Modupe Odusami et al. entitled "Explainable Deep-Learning-Based Diagnosis of Alzheimer's Disease Using Multimodal Input Fusion of PET and MRI Images" explores multimodal input fusion in PET and MRI imaging for Alzheimer's disease diagnosis [9]. Their proposed heuristic early feature fusion framework improved upon the unimodal models' diagnostic accuracy. The study has shown that data fusion can offer advantages for multimodal applications in neuroimaging. Álvaro S. Hervella et al discuss the use of a self-supervised learning scheme for diabetic retinopathy classification in "Multimodal Image Encoding Pre-Training for Diabetic Retinopathy Grading" [10]. The method takes advantage of the use of multiple retinal imaging modalities to increase grading accuracy and underscores the promise of pre-training techniques for multimodal healthcare applications. Finally, in "ContIG: Self-Supervised Multimodal Contrastive Learning for Medical Imaging with Genetics," Aiham Taleb et al. [12] take a step forward to investigate deep learning in the fusion of medical data. In this paper, the authors develop a contrastive learning

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 1055- 1059
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0151

framework that combines genetic and imaging data for better disease prediction. This study presents a new approach to multimodal heterogeneous datasets and their treatment in improving health care decision-making. These studies cumulatively establish the growing trend of multimodal data integration in healthcare. Since traditional machine learning models used only unimodal data, there is growing importance of multimodal approaches for classification, prediction, and decision-making in disease scenarios, which is shown to be possible through the application of recent advancements in deep learning and XAI. Areas of active research include complexities involved in data fusion, feature selection, and model interpretability. Addressing these and other challenges will open the road for much more robust, reliable, and interpretable AI-driven healthcare solutions.

3. Methodology

3.1 Data collection

The research utilizes multimodal datasets including MRI, CT, PET imaging and other medical images both from public resources and proprietary databases. Different classification tasks including disease diagnosis, image watermarking and data fusion have received sample labeling from this group.

3.2 Data processing

The input images require preprocessing steps before use to make them functional for medical applications. Image preprocessing includes normalization that achieves standardization of pixel values while using DnCNN algorithms for denoising purposes. Principal Component Analysis (PCA) and other techniques perform dimensionality reduction for feature extraction.

3.3 Multimodal Data Fusion

A set of different methods exists for combining various data modalities under Multimodal Data Fusion. The first type of fusion merges extracted features between different modalities in advance of the classification process while the second approach combines independent results from different models that trained separately on unique modalities. Better feature representation occurs through transform-based fusion by implementing methods that include NSST and Fr-DTCWT in addition to RSVD.

3.4 Model Selection and Training

Multiple deep learning structures operate together for assessment as well as classification tasks. CNNs function primarily to extract features from medical images. The model integrates different information sources using attention systems alongside transformers. The system receives Explainable Artificial Intelligence (XAI) techniques with Concept Activation Vectors (CAVs) and Concept Localization Maps (CLMs) which enable better interpretation capabilities.

3.5 Experimental Setup

The training process of models happens on NVIDIA GPUs with TensorFlow and PyTorch frameworks running on their top-performance platforms. Adam optimization and scheduled parameters of Stochastic Gradient Descent (SGD) serve as performance optimization techniques. The results from learning models rely on different loss functions that include cross-entropy for classification and MSE for regression. The model becomes more resilient through applications of data augmentation techniques that involve random cropping and flipping and rotation methods.

3.6 Evaluation Metrics

Different evaluation criteria exist for assessing performance outcomes. Performance evaluation for classification tasks utilizes precision, recall, accuracy together with F1-score as measurement criteria. The performance metrics for image quality assessment include PSNR and SSIM. The models gain better interpretability through the implementation of explain ability measures SHAP values and Grad-CAM. Medical data integrity depends on security and robustness techniques to ensure its safety. NSCT and RSVD create secure digital marks by which watermarking methods function. The evaluation of adversarial robustness happens by running models toward adversarial attacks which helps confirm their worth in actual medical practices.

3.7 Statistical Analysis

Statistical validity requires ablation studies together with ANOVA and t-tests for determining statistical significance. The testing procedures enable model-to-model and approach-assessment comparisons resulting in research conclusions that are documented

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 1055- 1059
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0151

with statistically sound results. Deployment aspects of the system rely on using light models to process patient data in real-time through edge and cloud deployment services for telehealth practice applications. The organization maintains full adherence to HIPAA and GDPR data privacy requirements for legal and ethical procedures in handling patient data.

4. Results and Discussion

The experimental results indicate that the developed multimodal fusion techniques substantially improve both model reliabilities along with classification precision. All three fusion methods at feature-level and decision-level and transform-based levels contribute to enhanced performance compared to single-mode methods. The evaluation results confirm the success of the technique through high precision measurements and excellent recall and F1-score performance metrics. Both watermarking and adversarial robustness methods operate together as security mechanisms that protect data integrity and establish trustworthiness of the model implementation. The methodology demonstrates superior performance by integrating multiple data sources than traditional unimodal approaches because of their information diversity. The sophisticated data structures manage effectively due to the capabilities of deep learning algorithms particularly CNNs along with Transformers. Model interpretability becomes more achievable through explainable AI integration because it provides end users with insight into the decision-making algorithms. The positive findings need further consideration of relevant obstacles. Real-time systems currently face a challenge with too high a computational requirement from deep learning-based multimodal models. The availability of different multimodal datasets remains a substantial challenge that requires additional transfer learning and data augmentation methods for proper solutions. The following research needs to focus on developing computational capabilities and exploring performing fusion approaches for better outcome achievement. Table 1 shows the table demonstrate on review of studies using classical machine learning for multimodal classification [11-15].

.

Table 1 The Table Demonstrate on Review of Studies Using Classical Machine Learning for Multimodal Classification

The table demonstrate on review of studies using classical machine learning for multimodal classification					
Reference	Feature Extraction	Fusion Architecture	Primary Learner	Final Classifier	Modalities
Jafarian et al. [47]	CNN	early	CNN	FCNN	signal
Cheng et al. [61]	2D CNN, 1D CNN	early	2D CNN, 1D CNN	FCNN	signal
Jiang et al. [62]	CNN	early	CNN	FCNN	signal
Meng et al. [66]	CNN	early	CNN	SVM	signal
Wu et al. [67]	CNN	early	CNN	FCNN	signal
Yan et al. [68]	CNN	early	CNN	FCNN	signal
Jacenków et al. [20]	ResNet-50, BERT	late	Transformer	Multi-Layer Perceptron	image, text
Additional Study [48]	Transformer, CNN	hybrid	Multimodal Model	Softmax Classifier	image, text, metadata
Odusami et al. [33]	ResNet18	early	Deep Learning Model	Modified ResNet18	PET, MRI

Conclusion

Multimodal classification proves beneficial for medical imaging diagnosis solutions according to this research. A combination of deep learning models together with advanced fusion techniques forms an accurate diagnosis system that also secures patient data processing. By implementing explainable AI methods, the proposed approach achieves both interpretability and high-performance levels.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 1055- 1059
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0151

Demonstrations next in line will tackle how to increase the efficiency of the model and develop real-time deployment methods for telehealth [16].

References

[1]. Sleeman, W. C., Kapoor, R., & Ghosh, P., Multimodal Classification: Current Landscape, Taxonomy, and Future Directions, ACM Computing Surveys, Vol. 55 (2022), No. 7, Article 150.

[2]. Singh, O. P., Singh, A. K., & Zhou, H., Multimodal Fusion-Based Image Hiding Algorithm for Secure Healthcare System, IEEE Intelligent Systems, 2022.

[3]. Singh, K. N., Singh, O. P., & Singh, A. K. (2024). WatMIF: Multimodal Medical Image Fusion-Based Watermarking for Telehealth Applications. Cognitive Computation, 16(1947–1963).

[4]. Jacenków, G., O'Neil, A. Q., & Tsaftaris, S. A. (2024). Indication as Prior Knowledge for Multimodal Disease Classification in Chest Radiographs with Transformers. The University of Edinburgh, Canon Medical Research Europe, The Alan Turing Institute.

[5]. Khalid, F. I., Makhtar, M., Rosly, R., Hamzah, W. M. A. F. W., Sambas, A., & El-Ebiary, Y. A. B. (2024). Deep Neural Ensemble Classification for COVID-19 Dataset. Nanotechnology Perceptions, 20(S14), 584-598.

[6]. Cai, G., Zhu, Y., Wu, Y., Jiang, X., Ye, J., & Yang, D. (2023). A Multimodal Transformer to Fuse Images and Metadata for Skin Disease Classification. The Visual Computer, 39(2781–2793).

[7]. Mallya, M., & Hamarneh, G. (2022). Deep Multimodal Guidance for Medical Image Classification. Lecture Notes in Computer Science. Springer.

[8]. Zhang, N., Wu, H., Zhu, H., Deng, Y., & Han, X. (2022). Tomato Disease Classification and Identification Method Based on Multimodal Fusion Deep Learning. Agriculture, 12(2014).

[9]. Odusami, M., Maskeliūnas, R., Damaševičius, R., & Misra, S. (2023). Explainable Deep-Learning-Based Diagnosis of Alzheimer's Disease Using Multimodal Input Fusion of PET and MRI Images. Journal of Medical and Biological Engineering.

[10]. Hervella, Á. S., Rouco, J., Novo, J., & Ortega, M. (2022). Multimodal Image Encoding Pre-Training for Diabetic Retinopathy Grading. Computers in Biology and Medicine, 143(105302).

[11]. Lucieri, A., Bajwa, M. N., Braun, S. A., Malik, M. I., Dengel, A., & Ahmed, S. (2022). ExAID: A Multimodal Explanation Framework for Computer-Aided Diagnosis of Skin Lesions. arXiv:2201.01249.

[12]. Taleb, A., Kirchler, M., Monti, R., & Lippert, C. (2022). ContIG: Self-Supervised Multimodal Contrastive Learning for Medical Imaging with Genetics. Hasso Plattner Institute for Digital Engineering, University of Potsdam, Germany.

[13]. Dimitri, G. M., Spasov, S., Duggento, A., Passamonti, L., Lio, P., & Toschi, N., Multimodal Image Fusion via Deep Generative Models, University of Cambridge, 2022.

[14]. Kumar, S., & Sharma, S., An Improved Deep Learning Framework for Multimodal Medical Data Analysis, Big Data and Cognitive Computing, Vol. 8 (2024), Article 125.

[15]. Li, Y., Daho, M. E. H., Conze, P. H., Zeghlache, R., Le Boité, H., Tadayonif, R., Cochener, B., Lamard, M., & Quellec, G., A Review of Deep Learning-Based Information Fusion Techniques for Multimodal Medical Image Classification, arXiv preprint, 2024.

[16]. Zhang, C., Chu, X., Ma, L., Zhu, Y., Wang, Y., Wang, J., & Zhao, J., M3Care: Learning with Missing Modalities in Multimodal Healthcare Data, Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022.