National Conference on Behavioural Economics and Intelligent Decision Systems for Climate Change and Sustainable Development
International Journal of Healthcare Education & Medical Informatics
Volume 13, Issue 1&2 - 2026

Review Article

A Review of Deep Learning Fusion based Multimodal for Disease Diagnosis and Classification

Baljit Kaur[1], Navreet Kaur[2], Sunaina[3], Priya Thakur[4]

[1,2,3,4]Department of Computer Science and Engineering, DAV University, Jalandhar, Punjab, India
DOI: https://doi.org/10.24321/2455.9199.202605

I N F O

Corresponding Author:
Baljit Kaur, Department of Computer Science and Engineering, DAV University, Jalandhar, Punjab, India
E-mail Id:
kaurkajalbaljit@gmail.com
How to cite this article:
Kaur B, Kaur N, Sunaina, Thakur P. A Review of Deep Learning Fusion based Multimodal for Disease Diagnosis and Classification. J. HealthCare Edu. & Med. Inform. 2026;13(1&2):229-233.

Date of Submission: 2025-10-04
Date of Acceptance: 2025-10-29

A B S T R A C T

Deep Fusion multimodal learning has become an effective method for early disease diagnosis by blending varied sources of data, including medical images, clinical data, and genomic data. The reason for this method is to break the restrictions of unimodal systems that do not cover intricate cross-domain relations in patient data. New approaches utilise Convolutional Neural Networks (CNNs) for feature extraction, Vision Transformers (ViTs) to capture global attention, and self-supervised multimodal transformers to facilitate better cross-modal representation. These advances make it possible to detect conditions such as Alzheimer's and tumour growth at earlier stages and with higher precision. Yet significant challenges lie in data standardisation, model interpretability, and clinical validation in real-world settings. Future work can focus on developing more interpretable, scalable, and generalisable fusion architectures for accurate early disease prediction.

Introduction

Deep Multimodal Fusion Multimodal fusion learning demonstrates multimodal fusion attentiveness in the field of medical imaging for core areas like magnetic resonance imaging (MRI) and computed tomography. magnetic resonance imaging computed tomography (CT) and capsule endoscopy (CE), Mammographic image Computed tomography and mammography are used for the diagnosis and classification of various diseases. Various computerised treatments based on machines Computerised machines, or computer vision machines, have been utilised by most researchers for automatically identifying numerous protruding lesions in WCE images. It begins from the step of data or image pre-processing. Images are processed or enhanced in this step and then forwarded to the subsequent step for image segmentation. Utilised segmentation. Here are several techniques that are used in image segmentation: K-means and uniform app segmentation. the uniform roach, normal distribution, the uniform normal distribution, saliency, normal distribution, saliency, and saliency-based techniques. saliency. The segmented images now proceed further for feature extraction in the subsequent step. Then multimodals are employed for the fusion of the data. Fusion-based methods have recently provided fruitful results in the area of CV. These methods have also been successfully employed in medical image processing.[4,7,10]

The deep fusion learning-based method usually demands a large amount of data and massive computational power to train a model. Deep fusion learning is among the prevailing and probable methods in computer vision. Deep learning has also emerged to play a meaningful role in self-guiding medical diagnosis in the last few years.[9,5]

Various techniques based on deep fusion learning have been proposed to emphasise and diagnose several diseases.

Although imaging methods alone are not enough to confirm the presence of a disease diagnosis, they provide a basis value to diagnose the disease along with other diagnostic techniques like blood tests, clinical histories, hand-designed features, etc.

Thus, the integration of various modalities and techniques using the assistance of deep fusion learning can definitely mechanise the diagnosis process and enhance the quality and precision of the disease recognition. Although many studies have made efforts to design machine-learning modalities for the medical diagnosis of various diseases.[7,8]

Materials and Methods

A well-organised, arranged, qualified method and modality is a crucial element in the favourable outcome of high-precision prognosis of diseases.[7] Deep Fusion-based learning (DL) is utilised in medical imaging as a computer vision technology. The diagnostic methods must improve medical image analysis for disease detection. A well-prepared, structured diagnosis enables pre-processing, mechanised planning to leverage the appropriate intervention technology.[2]

Researchers have proposed many deep learning methods for medical image recognition and diagnosis, such as CNN and other models. CNN is a concatenated model in which a multiple number of layers are utilised to examine thousands of images in the dataset and extract the correct features from them to identify infections, tumours, cancer stages, etc. CNN models possess good skills to capture precise features that the human eye cannot see.[6]

The interpretation of multi-model deep learning approaches has remarkable progress in technology by improving diagnostic accuracy and efficiency. The contributions of various models, such as VGG-16, VGG-19, ResNet, AlexNet, and GoogleNet, are put into images, MRI, CT, ultrasound, handcrafted features, mammographic images, and chest X-ray datasets for multiclass classification purposes.[3,1,6]

The good planning and organised structure of methods and techniques used in any models will give the optimal results

This section presents the flowchart, which is illustrating the multimodal deep fusion pipeline for disease classification, showing the step-by-step process from acquiring the input image and preprocessing to the use of multiple applied modalities, feature extraction, and feature fusion for generating the end diagnostic output, as shown in Figure 1.

Image Data

The data used consists of: clinical data, hand crafted features, mammographic images, another images data such as echocardiography, CT MRI, and X-ray etc.[1,2,5-7]

Image Preprocessing

The image data are enhanced through a series of pre-processing steps. These images are optimised in terms of exposure, contrast, and noise removal, and then the diversity and size of the training samples are increased using data augmentation techniques such as rotation, scaling, and random cropping.[2,3]

Feature Extractions

Post image preprocessing, the models VGG-16, VGG-19, ResNet, AlexNet, GoogleNet, Vision Transformers, etc. are employed to get divergent features from the preprocessed images. VGG-16 and VGG-19 are employed to obtain global structures, ResNet is employed to operate on high-level textures, AlexNet concentrates on extracting fine-grained edges, GoogleNet delivers multi-scale features, and vision transformers operate on attention mechanisms. These features are aggregated into a holistic feature vector.[1,3,8,9]

Different Models and Algorithm for Feature Extraction

- **Convolution-Vision Transformers:** 'CviT' is a convolution and transformer networks to identify movement classification, worked along with anattention-based transformer model.[8]
- **Inception-v3:** Inception-V3 focuses on feature extraction by dimensionality reduction without losing the model efficiency.[8]
- **Resnet50**: It is employed for extracting features from the Inception-V3 model, while the ResNet-50 emphasises low-level features as well as use the residual connections within the architecture.[8]
- **Resnet101 :** This was inspired by VGG19 pre-trained network and one of the deepest convolutional neural network (CNN) network.
- **Artificial Neural Network :** Enabling the model to learn complex patterns within the feature selection.[4]
- **PCA:** Principal Component Analysis (PCA) is also applied to dimensionality reduction of the fused feature vector. This process eliminates redundant information while retaining the most significant features.[4]
- **Ensemble Feature Selection (EAMSO Approach)**: The EAMSO approach integrates feature selection outputs from Artificial Ecosystem-based Optimisation (AEO), Monarch Butterfly Optimisation (MBO), and the Seagull Algorithm.[4]
- **VGG-16andVGG-19**: Extracted global structural features through their deeper convolution layers.[1]

Feature Fusion

After feature extraction, feature fusion is performed to combine these extracted features into a unified repre-

ISSN: 2455-9199
DOI: https://doi.org/10.24321/2455.9199.202605

sentation. The concatenated feature vector combines the strengths of each model, capturing a broader range of optimal information.

Literature Survey

After feature extraction, feature fusion is performed to combine these extracted features into a unified representation. The concatenated feature vector combines the st Many researchers have proposed various methods, models and techniques. Despite these methodologies, the literature gives out considerable gaps in multi-model approaches, particularly concerning their efficiency in classification recognitions.

[1]Xiangchun Yu and Wei Pang introduced Model1 based on VGG16 and Model2 based on VGG19. The VGG16 model operates to combine the deep features to distinguish the normal and tumour ROI patches. The VGG19 model worksto combine the cross-channel deep features and to obtain the final prediction by executing the majority voting of all patches' predictions of a single ROI. In order to effectively target the recognition of abnormalities in mammographic images. But it needs high computational resources.

[2]Tao Yu1 & Ke Yue Chen2 proposed a hybrid multimodal data fusion that integrates machine learning techniques with medical image analysis to improve the identification of cardiovascular diseases that integrates clinical data and echocardiogram images (segmentation via FCN/U-Net, normalisation, noise reduction). Feature extraction using MLP+Resnet-50 CNN and feature fusion (MLP+CNN features combined). Non-invasive diagnosis can reduce reliance on costly/risky angiography, and high accuracy can be achieved. Limited to retrospective datasets – may need real-time clinical validation.

[3]Kumar, R., Pan, C. T., Lin, Y. M., Yow-Ling, S., and Chung, T. S., introduced the EMDL framework that gives out an efficient outcome for accurate pulmonary disease diagnosis. Integrates five pre-trained CNNs: VGG-16, VGG-19, ResNet, AlexNet, GoogleNet… … Advanced Image Processing (Histogram Equalisation & ICEA); Feature Selection and Optimisation Pipeline; PCA; Final Classification using SVM. Faster and more accurate, handles multiclass classification, and balanced dataset (avoids bias). High computational cost, Complex pipeline (multimodel + multistep optimisation -> longer training time)

[4]Dhivyaa, C. R., Nithya, K., Kumar, C. S., & Sudhakar, R. introduced The proposed Trio FusionNet model for tuberculosis detection in chest X-ray images combines deep learningarchitectures, including ResNet-50, Inception V3, and EfficientNet-B4, for efficient feature extraction and An ensemble of AMS optimisation model approaches is developed for selecting the significant features from the extracted feature sets. Robust feature selection via EAMSO ensemble reduces redundancy. Higher computational cost; dataset / practical limitations.

[5]Fati, S. M., Senan, E. M., & ElHakim, N. (2022) introduced a deep and hybrid learning technique for early detection of tuberculosis based on X-ray images using feature fusion which integrates Hybrid CNN + SVM and the ANN Classifier with Feature Fusion. Handles class imbalance with augmentation; the ANN fusion system generalises better across datasets. Dependent on publicly available datasets (may not fully generalise to clinical settings).

[6]Ahmed, I. A., Senan, E. M., & Shatnawi, H. S. A. Hybrid models for endoscopy image analysis for early detection of gastrointestinal diseases are based on fused features, which include hybrid CNN–FFNN and CNN–XGBoost models. Feature fusion captures complementary patterns from multiple CNNs. Both FFNN and XGBoost provide robust classification with reduced redundancy (via PCA). Requires computationally expensive CNN training and fusion; endoscopy images are still subject to artefacts.

[7]Khan, M. A., Sarfraz, M. S., Alhaisoni, M., Albesher, A. A., Wang, S., & Ashraf, I. proposed StomachNet: Optimal deep learning features fusion. Hybrid CNN + SVM; ANN classifier with feature fusion for stomach abnormalities classification. improve accuracy, consume less time and remove the problem of overfitting. Their failure to extract good features in the classification phase.

[8]S., Ivanova, O., Melyokhin, A., & Tiwari, P. Proposed Intermediate Fusion Model ,Late Fusion Model Deep-learning-enabled multimodal data fusion for lung disease classification. Learn the mappings between heterogeneous data to benefit the decision process.

[9]Dutta, P., Sathi, K. A., Hossain, M. A., & Dewan, M. A. A. (2023). Conv-ViT: a convolution and vision transformer-based hybrid feature extraction method for retinal disease detection. This model is helpful in macular degeneration grading. Despite its high computational complexity.

[10]Hans, R., & Kaur, H. (2020). Hybrid binary Sine Cosine Algorithm and Ant Lion Optimisation (SCALO) approaches for the feature selection problem. Best suited for removing the inappropriate features.

ISSN: 2455-9199
DOI: https://doi.org/10.24321/2455.9199.202605

Table 1.Comparative Analysis of Deep Learning Fusion based Multimodals

REF. NO	TITLE	AUTHOR & YEAR	DATASET NAME	PROPOSED MODELS	COMPARISION WITH ANOTHERS MODELS	KEY FINDINGS
[1]	Models of VGG for Mammograph	Xiangchun Yu & Wei Pang…2020	Mammographic ROI pathches	Model1-VGG16, Model 2 –VGG 19	More accurate tumor vs. normal classification; improved ROI patch-level identification	Fuses deep and cross-channel features; suitable for mammogram abnormality detection.
[2]	Hybrid Fusion for CVD Detection	Tao Yu & Ke Yue Chen……2025	Clinical and Echocardiogram data	Hybrid Multimodal ML + CNN(Resnet50 +MLP)	Enhances over isolated imaging/ clinical models; enhances diagnostic accuracy	Non-invasive CVD detection; combines multimodal information effectively.
[3]	EMDL Model for Pulmonary Disorders	Kumar, R.et al……2025	Pulmonary X-ray Dataset	EMDL (VGG16 +VGG19 +ResNet +AlexNet +GoogleNet + SVM)	Faster & more accurate than a single CNN model	Supports multiclass scenarios; balanced data set; pipeline optimised.
[4]	TrioFusionNet for Tuberculosis Detection	Dhivyaa,C.R.et al……2024	Chest X-ray Dataset	TrioFusionNet (ResNet50 + InceptionV3 + EfficientNet-B4 + AMSO)	You outperform standalone CNNs and ensemble baselines	Strong feature extraction & selection; enhanced TB detection
[5]	Hybrid CNN-SVM for TB Diagnosis	Fati,S.M. et al….2022	Public TB X-ray Dataset	Hybrid CNN + SVM + ANN Fusion	Improved generalization and accuracy over CNN alone	Deals with class imbalance; cross-dataset generalization
[6]	Hybrid Models for GI Detection	Ahmed, I.A. et al….2023	Endoscopy Image Dataset	Hybrid CNN–FFNN, CNN–XGBoost	Enhances robustness and accuracy compared to sole CNN	Feature fusion detects complementary CNN patterns
[7]	StomachNet: Deep Feature Fusion	Khan, M.A. et al….2020	Stomach Image Dataset	StomachNet (Hybrid CNN + SVM + ANN Fusion)	Increases accuracy, decreases overfitting, and is faster than the regular CNNs	Optimal feature fusion enhances classification

ISSN: 2455-9199
DOI: https://doi.org/10.24321/2455.9199.202605

[8]	Fusion Models for Lung Diseases	S.,Ivanova, O.,Melyokhin,A.,and Tiwari,P.....2023	Lung Imaging Dataset	Intermediate & Late Fusion Models	Improves decision-making compared to single-modality CNNs	Trains on multimodal data; better classification
[9]	Conv-ViT for Retinal Diseases	Dutta,P.et al.....2023	Retinal Fundus Images	Conv-ViT (CNN + Vision Transformer)	Better than CNN-alone approaches in macular degeneration grading	Detection accuracy improves through hybrid feature extraction
[10]	SCALO for Feature Selection	Hans,R.& Kaur,N.....2020	Benchmark Feature Datasets	Hybrid Sine Cosine + Ant Lion Optimization	Outperforms basic metaheuristics for feature reduction	Efficiently discards redundant features; enhances classification

Future Directions

The Review also significantly contributes to improving the diagnostic methodologies. Results showed that the proposed methods outperform advanced approaches and it also shows better accuracy. The Continuation work is to further investigate the deep fusion learning which can distinguish the contribution of various branches. The model and their mechanism 's robustness can be improved and fortify by fusing various levels of patterns by different weights.

Follow-up work will include Real time deployment in hospitals, further modalities (wearable sensors, ECG) and Long term Goal predict disease progression years ahead of time using continuous multimodal monitoring.

References

1. Yu, X., Pang, W., Xu, Q., & Liang, M. (2020). Mammographic image classification with deep fusion learning. *Scientific Reports*, *10*(1), 14361.
2. Yu, T., Chen, K. Enhancing cardiac disease detection via a fusion of machine learning and medical imaging. *Sci Rep* **15**, 26269 (2025). https://doi.org/10.1038/s41598-025-12030-6
3. Kumar, R., Pan, C. T., Lin, Y. M., Yow-Ling, S., Chung, T. S., & Janesha, U. G. S. (2025). Enhanced Multi-Model Deep Learning for Rapid and Precise Diagnosis of Pulmonary Diseases Using Chest X-Ray Imaging. *Diagnostics*, *15*(3), 248.
4. Dhivyaa, C. R., Nithya, K., Kumar, C. S., & Sudhakar, R. (2024). Explainable Model of Fusion Network With Enhanced Optimization Approach for Tuberculosis Diagnosis. *IEEE Access*.
5. Fati, Suliman Mohamed, Ebrahim Mohammed Senan, and Narmine ElHakim. "Deep and hybrid learning technique for early detection of tuberculosis based on X-ray images using feature fusion." *Applied Sciences* 12.14 (2022): 7092.
6. Ahmed, I. A., Senan, E. M., & Shatnawi, H. S. A. (2023). Hybrid models for endoscopy image analysis for early detection of gastrointestinal diseases based on fused features. *Diagnostics*, *13*(10), 1758.
7. Khan, M. A., Sarfraz, M. S., Alhaisoni, M., Albesher, A. A., Wang, S., & Ashraf, I. (2020). StomachNet: Optimal deep learning features fusion for stomach abnormalities classification. *IEEE Access*, *8*, 197969-197981.
8. Kumar, S., Ivanova, O., Melyokhin, A., & Tiwari, P. (2023). Deep-learning-enabled multimodal data fusion for lung disease classification. *Informatics in Medicine Unlocked*, *42*, 101367.
9. Dutta, P., Sathi, K. A., Hossain, M. A., & Dewan, M. A. A. (2023). Conv-ViT: a convolution and vision transformer-based hybrid feature extraction method for retinal disease detection. *Journal of Imaging*, *9*(7), 140.
10. Hans, R., & Kaur, H. (2020). Hybrid binary Sine Cosine Algorithm and Ant Lion Optimization (SCALO) approaches for feature selection problem. *International Journal of Computational Materials Science and Engineering*, *9*(01), 1950021.

ISSN: 2455-9199
DOI: https://doi.org/10.24321/2455.9199.202605