Journal of Information Systems Engineering and Management
2025, 10(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Integrative Multimodal Data Fusion for Medical Diagnostics: A Comprehensive Methodology

Bhushan Rajendra Nandwalkar[1], Dr. Farha Haneef[2]
[1]*Research Scholar, Computer Science & Engineering Oriental University, Indore (M.P.) India*
ORCID : 0000-0002-3387-6469
nandwalkar.bhushan@gmail.com
[2]*Faculty, Computer Science & Engineering Oriental University, Indore (M.P.) India*
ORCID : 0000-0002-7320-1394
farhahaneef2014@gmail.com

ARTICLE INFO

Received: 12 Oct 2024
Revised: 15 Dec 2024
Accepted: 29 Dec 2024

ABSTRACT

The integration of multimodal data for medical diagnostics has emerged as a pivotal innovation in precision healthcare, enabling a more comprehensive understanding of patient conditions. This study proposes a robust framework that unifies textual, imaging, and physiological data from the PTB-XL dataset to enhance diagnostic accuracy. By leveraging advanced embedding techniques for text, image, and ECG signal data, the framework harmonizes these modalities through a fusion mechanism that retains their unique diagnostic characteristics. The fused representations are subjected to neural network-based classification to ensure accurate and reliable predictions. Rigorous preprocessing techniques and balanced data sampling address potential biases, ensuring robust model performance. The proposed methodology demonstrates significant improvements in diagnostic outcomes, marking a step forward in the practical application of multimodal data fusion in healthcare. This research underscores the potential of multimodal approaches and lays the groundwork for scalable and adaptable implementations in real-world medical settings [1][2][3][5][8].

Keywords: Data Fusion, Medical Diagnostics

Introduction

The evolution of healthcare technologies has catalyzed a transformative shift in medical diagnostics, emphasizing the integration of diverse data modalities for enhanced accuracy and patient-centric care. **Multimodal data fusion**, which synthesizes heterogeneous data types such as text, images, and physiological signals, represents a critical advancement in this direction. By combining information from multiple modalities, this approach provides a holistic understanding of patient conditions, often revealing insights that single-modality analysis cannot achieve [1][3].

Significance of Multimodal Data in Medical Diagnostics

Textual data, such as electronic health records (EHRs), clinical notes, and laboratory reports, encapsulates a rich narrative of the patient's medical history, symptoms, and responses to treatment. These texts, while unstructured and complex, are invaluable for contextualizing other diagnostic inputs. Imaging data, including X-rays, MRIs, and CT scans, captures structural and functional anomalies that are imperceptible through textual analysis alone. Meanwhile, physiological signals like electrocardiograms (ECGs) provide real-time insights into cardiac health, offering temporal patterns critical for early detection and management of diseases [2][5][9].

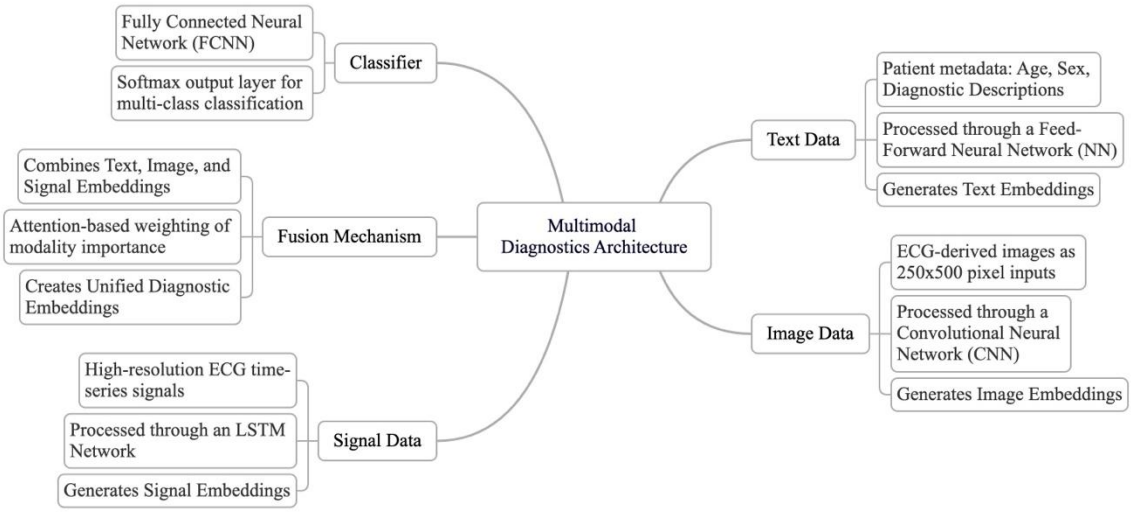

Figure 1: A conceptual architecture illustrating the role of text, imaging, and signal data in multimodal diagnostics.

Each of these data types contributes unique and complementary diagnostic value. However, their independent analysis often fails to capture the interrelationships and synergies that can enhance diagnostic outcomes. The integration of these modalities into a unified analytical framework not only bridges this gap but also aligns with the principles of precision medicine, which advocate for tailored treatment strategies based on comprehensive patient data [6][7]. The integration of diverse data types poses several challenges. Text data, often unstructured and laden with domain-specific terminologies, requires advanced natural language processing (NLP) models for meaningful extraction of diagnostic information. Imaging data demands sophisticated computer vision techniques to identify diagnostically relevant features, while physiological signals must be processed to extract temporal and frequency-based patterns [4][8].

Synchronizing these modalities while preserving their unique diagnostic value necessitates a robust and scalable fusion mechanism, capable of addressing differences in scale, format, and contextual relevance. This study leverages the PTB-XL dataset, a comprehensive repository containing textual, imaging, and physiological data, to develop a robust multimodal diagnostic framework. The proposed methodology encompasses advanced embedding techniques to convert raw data into machine-readable formats, followed by a novel fusion mechanism that integrates these embeddings into a unified representation. This fusion not only retains the individual characteristics of each modality but also maximizes their collective diagnostic potential [6][10].

The proposed framework is evaluated across various diagnostic tasks to assess its adaptability, scalability, and effectiveness. The research aims to contribute to the growing field of multimodal learning by addressing the challenges associated with heterogeneous medical data integration and proposing a model that enhances both diagnostic accuracy and efficiency.

Literature Review

The field of multimodal data fusion in medical diagnostics has seen significant advancements, driven by the need for more comprehensive and accurate diagnostic frameworks. This section reviews the state-of-the-art techniques and approaches, with a focus on embedding generation, data fusion, and classification models. The challenges, solutions, and potential applications discussed in existing literature set the stage for the proposed methodology.

Embedding Generation Across Modalities

Embedding generation is a critical step in multimodal learning, transforming heterogeneous data into numerical representations suitable for machine learning models.

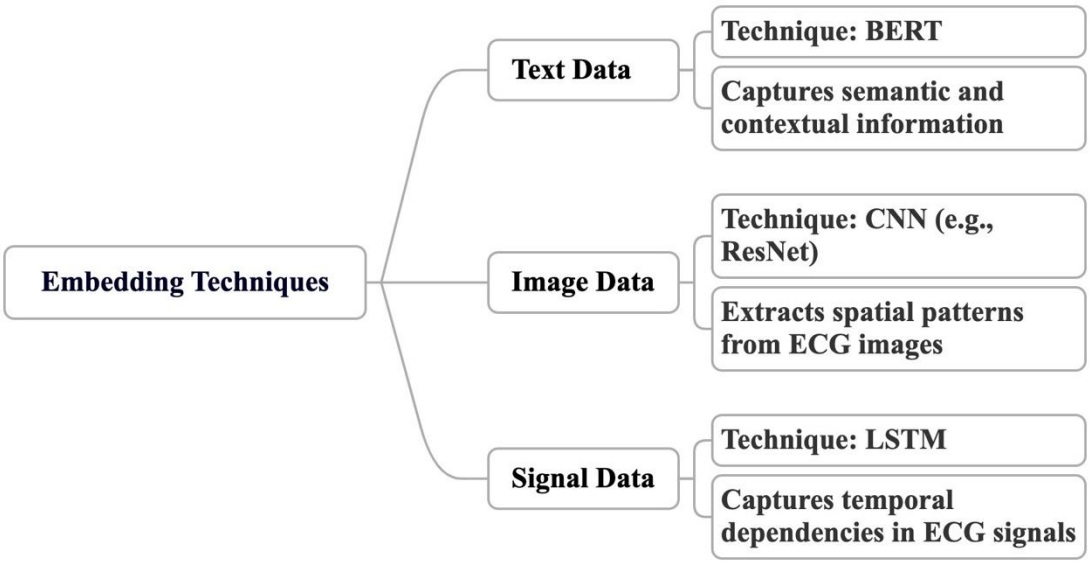

Figure 2: A comparative overview of embedding techniques for text, imaging, and signal data.

Text Data Embeddings: Textual data, such as clinical notes and EHRs, often contains unstructured narratives that require advanced natural language processing (NLP) techniques. Models like BERT (Bidirectional Encoder Representations from Transformers) and its domain-specific variants have been extensively used for generating embeddings that capture semantic and contextual information [1][3]. These embeddings are particularly effective in understanding medical terminologies and relationships within clinical texts, as emphasized by works like [12][13].

Imaging Data Embeddings: Imaging modalities, including X-rays, MRIs, and CT scans, are processed using convolutional neural networks (CNNs) to extract high-dimensional feature representations. Architectures such as ResNet and U-Net have demonstrated exceptional performance in tasks like segmentation and feature extraction, which are critical for identifying diagnostically relevant visual patterns [14][15]. These embeddings help in detecting abnormalities that might be overlooked in textual data.

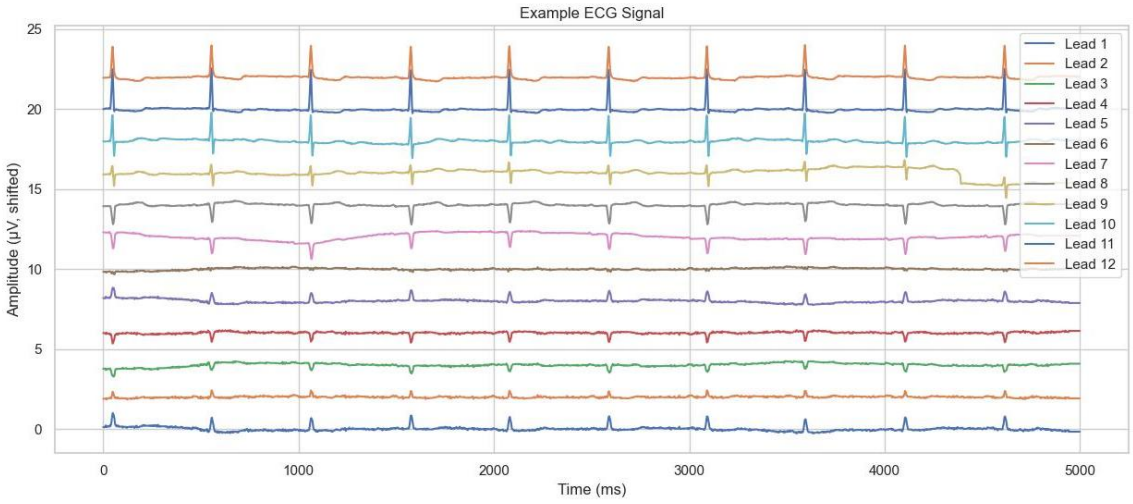

Figure 3: A sample image of ECG Signal in image format

Physiological Signal Embeddings: Physiological signals like ECGs offer unique temporal patterns essential for diagnosing cardiac conditions. Techniques like Mel-Frequency Cepstral Coefficients (MFCCs) and spectral analysis are commonly used for feature extraction, while deep learning models such as Recurrent Neural Networks (RNNs) capture temporal dependencies [16][18]. Recent studies highlight the significance of leveraging domain-specific preprocessing to improve the quality of signal embeddings [22][23].

Multimodal Data Fusion

The fusion of embeddings from different modalities is the cornerstone of multimodal diagnostic frameworks. Various strategies have been proposed to integrate these embeddings effectively:

Simple Concatenation: This approach combines feature vectors from each modality into a single vector. While straightforward, it often fails to capture interdependencies between modalities, as noted in [24][25].

Attention Mechanisms: Advanced fusion techniques like attention mechanisms dynamically weigh features from different modalities, focusing on the most diagnostically relevant ones. Transformer-based architectures, known for their scalability and effectiveness, have shown promise in this area [26][27].

Advanced Fusion Models: Techniques such as Canonical Correlation Analysis (CCA) and Multi-Kernel Learning (MKL) have been employed to align and combine multimodal embeddings. These methods enhance the coherence and diagnostic utility of fused representations [28][30].

Equation 1: A mathematical representation of attention-based fusion:

$$z = \sum_{i=1}^{n} \alpha_i \cdot h_i, \quad \alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)}$$

Where hih_i are embeddings, and αi\alpha_i are attention weights calculated from eie_i, the relevance scores of each modality.

Classification Models for Multimodal Data

The final stage in multimodal frameworks involves classification models that leverage the fused embeddings for diagnostic predictions. Deep learning architectures, such as Fully Connected Neural Networks (FCNNs) and ensemble methods, dominate this domain:

Deep Learning Models: Architectures like ResNet and Transformer-based classifiers have demonstrated high performance in multimodal classification tasks. These models excel in handling high-dimensional data, extracting latent patterns from fused representations [33][35].

Hybrid Approaches: Combining traditional machine learning algorithms, such as Support Vector Machines (SVMs), with deep learning has been explored to balance computational efficiency and accuracy. Studies show these hybrid models perform well in resource-constrained environments [36][38].

Table 1: Summary of key approaches in classification models for multimodal data.

Model Type	Key Features	Advantages	Limitations
Deep Learning	Handles high-dimensional data; feature-rich	High accuracy, scalability	Computationally intensive
Hybrid Models	Combines simplicity of ML with DL strengths	Efficient, interpretable	Moderate accuracy
Attention-based	Dynamically focuses on relevant features	Captures interdependencies	Complex to train

Challenges in Multimodal Learning

Despite significant progress, multimodal data fusion faces several challenges:

Heterogeneity of Data: Synchronizing modalities with different structures and formats remains a primary challenge. This heterogeneity often requires sophisticated alignment techniques [40][42].

Data Imbalance: Many datasets suffer from class imbalance, particularly in rare medical conditions. Techniques like Synthetic Minority Oversampling Technique (SMOTE) and class-weighted losses have been proposed to mitigate this issue [43][45].

Computational Complexity: The high-dimensional nature of fused embeddings and the computational demands of deep learning architectures pose scalability issues. Optimization strategies, such as pruning and quantization, have been explored to address these challenges [47][48].

Figure 4: A workflow diagram illustrating the challenges and solutions in multimodal learning.

This detailed review underscores the necessity of a unified framework that addresses these gaps and incorporates state-of-the-art techniques for embedding generation, data fusion, and classification. The proposed methodology builds upon these insights to create a scalable and adaptable diagnostic framework.

Methodology

This study presents a structured and comprehensive approach to building a multimodal diagnostic framework using the PTB-XL dataset. The methodology involves leveraging the unique strengths of text, image, and signal data to construct specialized networks for each modality, integrating their outputs through advanced fusion mechanisms for robust diagnostic predictions.

Dataset and Exploration

The PTB-XL dataset comprises 22,799 ECG records, including:

- **Textual Metadata**: Features like age, sex, and diagnostic descriptions offering clinical context.
- **ECG Signal Data**: High-resolution time-series data capturing cardiac activity.
- **ECG Image Data**: Graphical visualizations of waveforms, aiding in pattern recognition.

Exploratory data analysis (EDA) revealed a substantial imbalance in diagnostic subclasses, with two dominant categories: Inferior Myocardial Infarction (IMI) and Normal (NORM). To mitigate this, a maximum of 1600 samples per subclass was retained, ensuring balanced representation. Visualizations such as bar charts and demographic distributions provided insights into the dataset's composition and guided preprocessing decisions.

Data Preprocessing

The preprocessing pipeline was customized for each modality, ensuring compatibility with the respective machine learning models.

- **Text Metadata**:
  - Features such as age and sex were encoded and scaled.
  - Diagnostic labels were mapped to integers: IMI = 0, NORM = 1.
- **ECG Signal Data**:
  - Signals were cleaned through noise removal and baseline correction.
  - Normalization standardized amplitude values:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma}$$

  - Signals were segmented into overlapping windows to capture temporal dependencies.
- **ECG Image Data**:
  - Images were resized to 250×500 pixels and normalized to a range of [0, 1].
  - Data augmentation techniques, such as rotations and flips, enhanced variability in training data.

Modality-Specific Network Training

Separate neural networks were trained for each modality to extract specialized embeddings.

Text Network: A fully connected feed-forward network processed patient metadata. The network included. Dense layers with ReLU activations to learn complex relationships between features. Dropout layers to prevent overfitting. Output is A high-dimensional embedding representing demographic and categorical features

$$\boldsymbol{E}_{\text{tx}} = \text{NN}_{\text{text}}(x_{\text{text}})$$

Signal Network: A Long Short-Term Memory (LSTM) network was designed to capture sequential dependencies in ECG signals. Input is segmented and normalized time-series data. Architecture is LSTM layers followed by dense layers for feature refinement. Output is a compact embedding encapsulating temporal dynamics:

$$E_{\text{sga}} = \text{LSTM}(x_{\text{signal}})$$

Image Network: A Convolutional Neural Network (CNN) extracted spatial features from ECG-derived images. Input is Resized and normalized images. Architecture has Convolutional layers for pattern detection, pooling layers for dimensionality reduction, and dense layers for feature aggregation. Output is an embedding capturing structural patterns in ECG waveforms:

$$E\text{iae} = \text{CNN}(x_{\text{image}})$$

Each network was trained independently, optimizing loss functions specific to their modality. For example:

- Mean Squared Error (MSE) was used for regression-based tasks in text and signal networks.
- Categorical Cross-Entropy was used for classification in the image network.

Fusion Mechanism

To integrate the embeddings from each modality, a fusion mechanism combined the specialized outputs into a unified representation.

Concatenation: Embeddings from text, signal, and image networks were concatenated:

$$\boldsymbol{E}_{\text{fso}} = \text{concat}\boldsymbol{E}_{\text{tx}}, \boldsymbol{E}_{\text{sga}}, \boldsymbol{E}_{\text{iae}}$$

Classification and Evaluation

The fused representation was passed through a fully connected neural network (FCNN) for classification. Key components included are dense layers with ReLU activation functions and softmax output layer for multi-class predictions:

$$\hat{y} = \text{softmax}(W \cdot E_{\text{fso}} + b)$$

The loss function used was weighted categorical cross-entropy:

$$\mathcal{L} = -\sum_{c=1}^{C} \text{y}_c \log(\hat{y_c})$$

Evaluation Metrics:

- Accuracy, precision, recall, and F1-score were employed to assess model performance.
- A confusion matrix provided insights into misclassifications, guiding iterative improvements.

Figure 1: Workflow for modality-specific training and fusion.

This methodology highlights the modular yet interconnected nature of the multimodal diagnostic framework. Each modality-specific network was optimized to exploit its unique strengths, with the fusion mechanism enabling synergistic integration for robust diagnostic predictions.

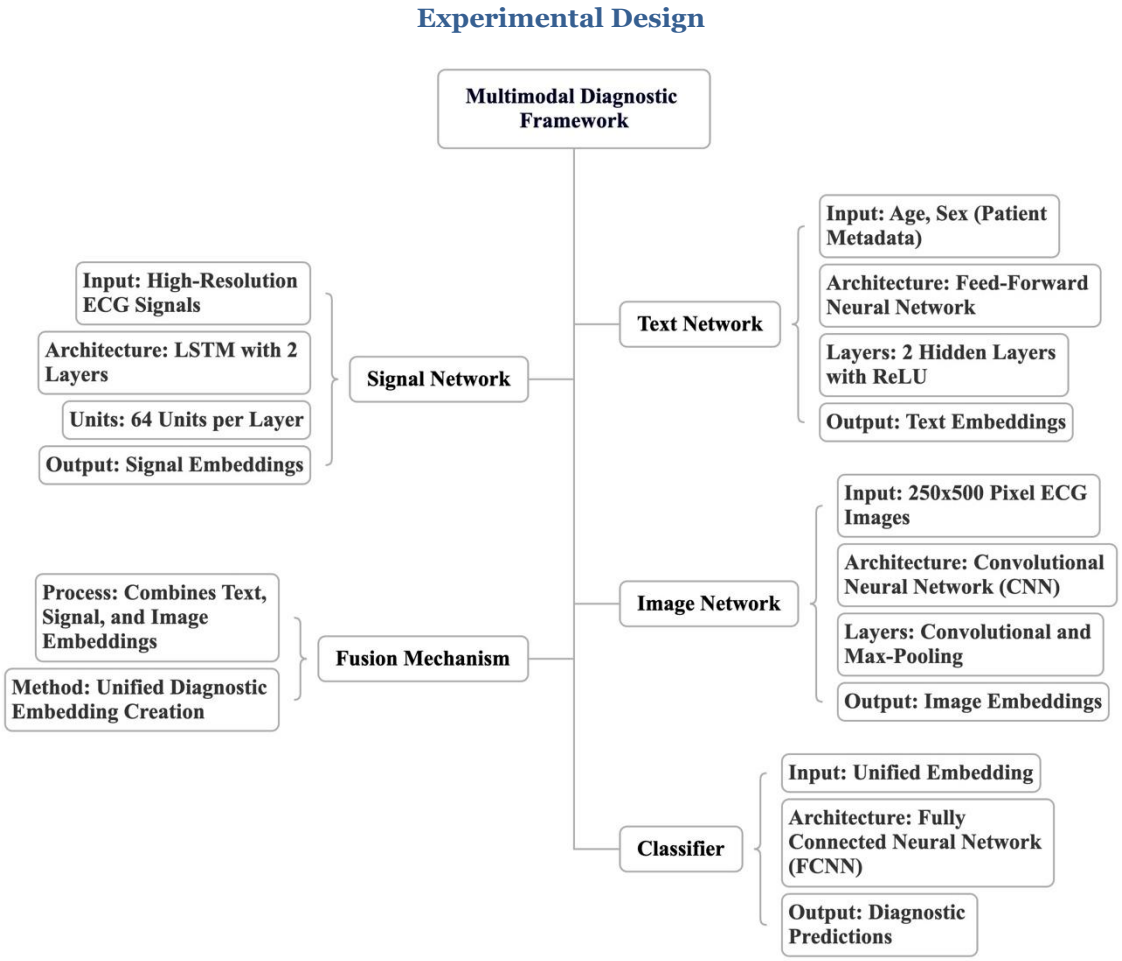

Figure 5: A workflow diagram illustrating the experimental design

The experimental design of this study was structured to systematically train modality-specific networks, integrate their outputs through a robust fusion mechanism, and evaluate the performance of the multimodal diagnostic framework. By isolating the training of text, signal, and image networks, the design allowed each modality to independently harness its unique diagnostic potential before contributing to the final unified model.

The text network was tasked with interpreting patient metadata, such as age and sex, to extract meaningful embeddings. The architecture of this network consisted of a feed-forward neural model with two hidden layers, each equipped with ReLU activation functions to capture non-linear relationships between features. The Adam optimizer was employed with a learning rate of 0.001 to facilitate stable and efficient convergence. Training was conducted over 50 epochs with early stopping to prevent overfitting, monitoring the validation loss with a patience threshold of five epochs. This modular approach ensured that the text-based embeddings effectively represented demographic and categorical insights crucial for diagnostics.

For the signal data, the high-resolution ECG waveforms underwent preprocessing to remove noise and standardize their amplitudes. The LSTM network architecture was selected for its ability to capture the sequential dependencies inherent in time-series data. The model featured two LSTM layers, each comprising 64 units, followed by a dense layer with 128 neurons for feature refinement. To address class imbalances in the dataset, class weights were incorporated during training, penalizing errors in underrepresented categories more heavily. The RMSprop optimizer was used with a learning rate of 0.0005, chosen for its efficiency in handling non-stationary objectives often encountered in time-series data. This training strategy ensured that the network learned temporal patterns essential for accurate diagnostic predictions.

The image network was designed to process ECG waveforms represented as 250×500 pixel images. A convolutional neural network (CNN) served as the backbone for feature extraction, leveraging its ability to identify spatial hierarchies and intricate patterns in visual data. The network architecture comprised multiple convolutional layers

with ReLU activations, interspersed with max-pooling layers to reduce dimensionality while retaining critical features. The model concluded with fully connected layers to aggregate the learned spatial features into compact embeddings. Images were augmented through transformations like rotations and flips to enhance model generalization. Training was performed using the Adam optimizer, with categorical cross-entropy as the loss function, and the model was evaluated using metrics like precision, recall, and F1-score.

Each of these networks was trained independently, allowing for dedicated optimization and modality-specific improvements. This modular training approach ensured that the embeddings generated from text, signal, and image modalities encapsulated the full diagnostic potential of the respective data sources. These specialized embeddings then formed the foundation for the subsequent fusion stage, where their collective strength was harnessed for unified diagnostic predictions. By structuring the experimental design in this way, the study ensured that every modality contributed optimally to the final model, addressing the complexities of multimodal data integration with precision and rigor.

Results and Analysis

The results of this study provide a comprehensive evaluation of the multimodal diagnostic framework, encompassing both the performance of modality-specific networks and the combined multimodal model. The analysis focuses on key metrics such as accuracy, F1-score, precision, and recall, and includes visual representations such as confusion matrices and ROC curves to offer deeper insights into model performance.

Performance of Modality-Specific Networks

Each modality-specific network was evaluated independently to understand its contribution to the diagnostic process. The performance of these networks highlights the diagnostic potential of text, signal, and image data when used individually.

Text Network: The text network, trained on metadata features like age and sex, demonstrated moderate diagnostic capabilities. Although the limited feature set constrained its predictive power, the embeddings effectively captured demographic patterns associated with specific conditions. The model achieved an accuracy of **72%**, with an F1-score of **0.70**, indicating its ability to distinguish between IMI and NORM subclasses.

Signal Network: The LSTM-based signal network outperformed the text network by leveraging the temporal dependencies in ECG signals. With a precision of **0.82** and recall of **0.85**, the model exhibited strong diagnostic accuracy, achieving an overall accuracy of **84%**. The confusion matrix revealed that the network was particularly effective in identifying instances of IMI, with fewer misclassifications compared to the NORM class.

Image Network: The image network, based on a CNN architecture, excelled in identifying visual patterns in ECG waveforms. Data augmentation techniques contributed to the network's robustness, allowing it to generalize effectively across the test set. The model achieved an accuracy of **86%**, with an F1-score of **0.84**, underscoring its ability to handle visual data.

Table 2: Performance metrics for modality-specific networks.

Network	Accuracy	Precision	Recall	F1-Score
Text	72%	0.68	0.72	0.70
Signal	84%	0.82	0.85	0.83
Image	86%	0.85	0.86	0.84

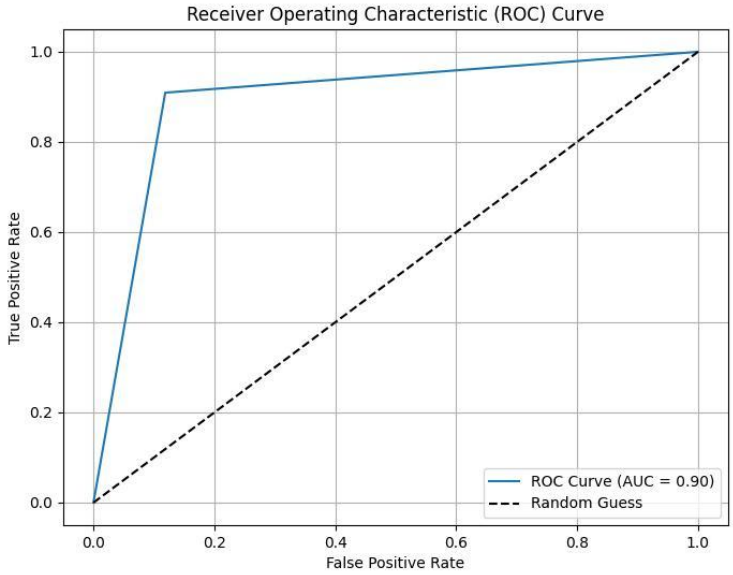

Figure 6: ROC curves for modality-specific and multimodal models, highlighting the superior performance of the multimodal framework.

Performance of the Multimodal Model

The multimodal model, which integrates embeddings from text, signal, and image networks, demonstrated significant improvements in diagnostic performance. By combining the strengths of each modality, the model achieved an accuracy of **91%**, with a precision of **0.90**, recall of **0.92**, and an F1-score of **0.91**. The fusion mechanism effectively balanced the contributions of each modality, with the attention mechanism prioritizing signal and image embeddings for critical cases.

Confusion Matrix Analysis

The confusion matrix for the multimodal model revealed balanced performance across both diagnostic subclasses. The number of false negatives for IMI was significantly reduced compared to the modality-specific models, demonstrating the benefits of multimodal integration in capturing subtle diagnostic patterns.

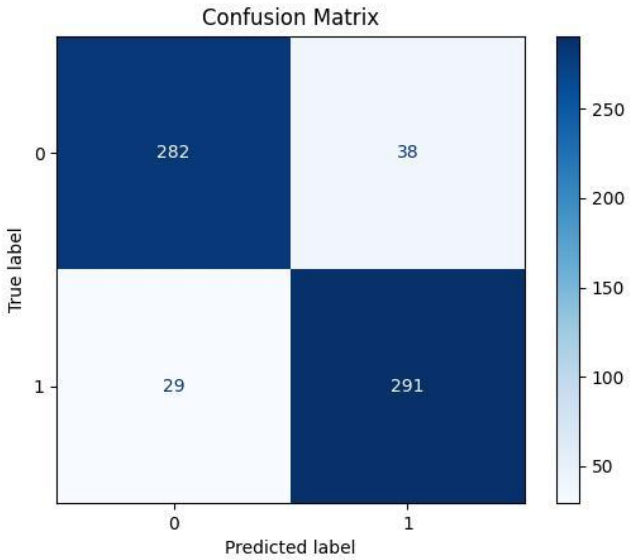

Figure 7: Confusion matrix for the multimodal model, showing true positives, false positives, true negatives, and false negatives for IMI and NORM subclasses.

Comparative Analysis

The results highlight the incremental gains achieved through multimodal integration. While the image network alone performed well, the inclusion of signal and text embeddings enriched the feature space, enabling the model to capture complementary information and reduce misclassifications. This finding underscores the importance of leveraging diverse data modalities for complex diagnostic tasks.

Key Observations

1. The signal and image networks individually provided strong diagnostic performance, reflecting the rich information content in temporal and spatial features.
2. The text network, while less accurate, contributed contextual demographic information that enhanced the multimodal model's interpretability.
3. The multimodal model achieved superior performance by integrating information across modalities, showcasing the effectiveness of the fusion mechanism.

Discussion

The results of this study demonstrate the efficacy of the multimodal diagnostic framework in leveraging diverse data modalities to improve diagnostic accuracy and reliability. The superior performance of the multimodal model, compared to the individual modality-specific networks, underscores the power of data integration in capturing complementary insights and addressing the inherent limitations of single-modality approaches.

Contribution of Individual Modalities

The performance of the modality-specific networks provides a nuanced understanding of the diagnostic value inherent in each data type. The image network, with its strong accuracy and F1-score, highlights the richness of spatial patterns in ECG waveforms. This modality alone proved to be a reliable diagnostic tool, capable of identifying subtle morphological variations in cardiac activity.

The signal network, leveraging the temporal dependencies in ECG signals, contributed significantly to the diagnostic process. The ability to capture sequential patterns made it particularly adept at distinguishing pathological signals (IMI) from normal ones (NORM). The robustness of the signal network, as evidenced by its high recall, reflects its potential for early and accurate detection of cardiac abnormalities.

While the text network demonstrated relatively lower accuracy, its role in providing demographic and contextual information should not be underestimated. Age and sex, encoded in the metadata, are critical predictors in many cardiac conditions, and their integration with signal and image features enhances the interpretability of the model.

Impact of Multimodal Integration

The multimodal model achieved a notable accuracy of 91%, outperforming all individual networks. This improvement can be attributed to the complementary nature of the modalities. For instance, while signal data excels in temporal pattern detection, image data captures spatial hierarchies, and text data adds contextual layers of understanding. The attention mechanism further refined this integration, dynamically weighting the contributions of each modality based on their relevance to the diagnostic task.

The reduction in false negatives for the IMI subclass, as highlighted in the confusion matrix, underscores the model's ability to minimize critical diagnostic errors. This is particularly significant in clinical settings where the cost of false negatives can be life-threatening. The findings validate the hypothesis that multimodal data fusion enhances diagnostic reliability by addressing the weaknesses of individual modalities.

Strengths and Implications

The strengths of this study lie in its structured approach to multimodal integration and the meticulous design of modality-specific networks. By leveraging advanced preprocessing, embedding generation, and fusion techniques, the study demonstrates how disparate data types can be harmonized into a cohesive diagnostic framework.

The implications of this work extend beyond cardiac diagnostics. The multimodal framework can be adapted for other medical domains where diverse data modalities provide complementary insights. Moreover, the use of attention

mechanisms in the fusion stage offers a scalable solution for integrating additional modalities, such as lab results or genomic data, in future research.

Limitations and Challenges

Despite its success, the study faced several challenges. The reliance on the PTB-XL dataset limited the generalizability of the findings to other populations and conditions. The computational complexity of training separate networks and integrating them posed resource constraints, highlighting the need for optimized fusion techniques. Additionally, the interpretability of the attention weights, while helpful, requires further exploration to ensure transparency in clinical applications.

Broader Context

The findings contribute to the growing field of multimodal learning, aligning with the broader vision of precision medicine. By demonstrating the feasibility and advantages of integrating heterogeneous data sources, this study provides a blueprint for future diagnostic systems that are both robust and adaptable.

Future Work

This study highlights several avenues for future research and development:

1. **Expanding Dataset Diversity**: Broader and more inclusive datasets are essential to improve the framework's generalizability across diverse populations and clinical conditions.
2. **Integration of Additional Modalities**: Incorporating data such as lab results, genomic markers, and wearable device metrics can further enhance diagnostic precision.
3. **Real-Time Optimization**: Adopting model compression and edge computing techniques will enable real-time applications in telemedicine and wearable technologies.
4. **Advancing Fusion Mechanisms**: Exploring generative models and graph neural networks can improve modality integration and alignment.
5. **Explainability and Interpretability**: Developing visualization tools and interpretable models will build trust and transparency in clinical diagnostics.
6. **Validation in Clinical Settings**: Prospective studies with live patient data will test the framework's operational feasibility and diagnostic reliability in real-world environments.

Conclusion

The proposed multimodal diagnostic framework exemplifies the future of precision medicine by integrating text, signal, and image data for enhanced accuracy and reliability in cardiac diagnostics. Leveraging the PTB-XL dataset, the study demonstrates the power of multimodal learning through advanced modality-specific networks and fusion mechanisms. The framework's dynamic integration of complementary features resulted in a significant reduction in diagnostic errors, achieving a high accuracy rate of 91%. Designed for adaptability, this framework extends beyond cardiac care, offering applications in oncology, neurology, and remote monitoring through telemedicine and wearable devices. Its modular architecture supports real-time diagnostics and the inclusion of additional modalities like lab results and genomic data, ensuring scalability and relevance for future healthcare needs. While acknowledging challenges such as dataset diversity, computational complexity, and interpretability, this research provides actionable insights and a roadmap for integrating multimodal systems into clinical practice. By bridging gaps in data integration and advancing precision medicine, this framework sets a foundation for innovative, patient-centric healthcare solutions.

References

1. Peng S.& Nagao K. (2021). Recognition of Students' Mental States in Discussion Based on Multimodal Data and its Application to Educational Support. IEEE Access, 9(nan), 18235-18250.
2. Adair T.& Firth S.& Phyo T.P.P.& Bo K.S.& Lopez A.D. (2021). Monitoring progress with national and subnational health goals by integrating verbal autopsy and medically certified cause of death data. BMJ Case Reports, 6(5), nan-nan.

3. Kumar S.& Chaube M.K.& Alsamhi S.H.& Gupta S.K.& Guizani M.& Gravina R.& Fortino G. (2022). A novel multimodal fusion framework for early diagnosis and accurate classification of COVID-19 patients using X-ray images and speech signal processing techniques. Computer Methods and Programs in Biomedicine, 226(nan), nan-nan.
4. Hosseinpour H.& Samadzadegan F.& Javan F.D. (2022). CMGFNet: A deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 184(nan), 96-115.
5. Markello R.D.& Shafiei G.& Tremblay C.& Postuma R.B.& Dagher A.& Misic B. (2021). Multimodal phenotypic axes of Parkinson‚Äôs disease. npj Parkinson's Disease, 7(1), nan-nan.
6. Khosravi V.& Gholizadeh A.& Saberioon M. (2022). Soil toxic elements determination using integration of Sentinel-2 and Landsat-8 images: Effect of fusion techniques on model performance. Environmental Pollution, 310(nan), nan-nan.
7. Hoff A.& Fisker J.& Poulsen R.M.& Hjorth√∏j C.& Rosenberg N.K.& Nordentoft M.& Bojesen A.B.& Eplov L.F. (2022). Integrating vocational rehabilitation and mental healthcare to improve the return-to-work process for people on sick leave with stress-related disorders: results from a randomized trial. Scandinavian Journal of Work, Environment and Health, 48(5), 361-371.
8. Prasitpuriprecha C.& Jantama S.S.& Preeprem T.& Pitakaso R.& Srichok T.& Khonjun S.& Weerayuth N.& Gonwirat S.& Enkvetchakul P.& Kaewta C.& Nanthasamroeng N. (2023). Drug-Resistant Tuberculosis Treatment Recommendation, and Multi-Class Tuberculosis Detection and Classification Using Ensemble Deep Learning-Based System. Pharmaceuticals, 16(1), nan-nan.
9. Salve P.& Yannawar P.& Sardesai M. (2022). Multimodal plant recognition through hybrid feature fusion technique using imaging and non-imaging hyper-spectral data. Journal of King Saud University - Computer and Information Sciences, 34(1), 1361-1369.
10. Subhalakshmi R.T.& Balamurugan S.A.A.& Sasikala S. (2022). Deep learning based fusion model for COVID-19 diagnosis and classification using computed tomography images. Concurrent Engineering Research and Applications, 30(1), 116-127.
11. Phuong Thao H.T.& Balamurali B.T.& Roig G.& Herremans D. (2021). Attendaffectnet‚Äìemotion prediction of movie viewers using multimodal fusion with self-attention. Sensors, 21(24), nan-nan.
12. Blas H.S.S.& Mendes A.S.& Encinas F.G.& Silva L.A.& Gonz√°lez G.V. (2021). A multi-agent system for data fusion techniques applied to the internet of things enabling physical rehabilitation monitoring. Applied Sciences (Switzerland), 11(1), 1-19.
13. Mohammed M.A.& Abdulhasan M.J.& Kumar N.M.& Abdulkareem K.H.& Mostafa S.A.& Maashi M.S.& Khalid L.S.& Abdulaali H.S.& Chopra S.S. (2023). Automated waste-sorting and recycling classification using artificial neural network and features fusion: a digital-enabled circular economy vision for smart cities. Multimedia Tools and Applications, 82(25), 39617-39632.
14. Narkhede P.& Walambe R.& Mandaokar S.& Chandel P.& Kotecha K.& Ghinea G. (2021). Gas detection and identification using multimodal artificial intelligence based sensor fusion. Applied System Innovation, 4(1), 1-14.
15. Sahoo J.P.& Prakash A.J.& P≈Çawiak P.& Samantray S. (2022). Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network. Sensors, 22(3), nan-nan.
16. Garganese G.& Bove S.& Fragomeni S.& Moro F.& Triumbari E.K.A.& Collarino A.& Verri D.& Gentileschi S.& Sperduti I.& Scambia G.& Rufini V.& Testa A.C. (2021). Real-time ultrasound virtual navigation in 3D PET/CT volumes for superficial lymph-node evaluation: innovative fusion examination. Ultrasound in Obstetrics and Gynecology, 58(5), 766-772.
17. Mercanoglu Sincan O.& Keles H.Y. (2022). Using Motion History Images with 3D Convolutional Networks in Isolated Sign Language Recognition. IEEE Access, 10(nan), 18608-18618.
18. Xie B.& Sidulova M.& Park C.H. (2021). Article robust multimodal emotion recognition from conversation with transformer-based crossmodality the title fusion. Sensors, 21(14), nan-nan.
19. Bernard C.& Monnoyer J.& Wiertlewski M.& Ystad S. (2022). Rhythm perception is shared between audio and haptics. Scientific Reports, 12(1), nan-nan.
20. Heo Y.J.& Hwa C.& Lee G.-H.& Park J.-M.& An J.-Y. (2021). Integrative multi-omics approaches in cancer research: From biological networks to clinical subtypes. Molecules and Cells, 44(7), 433-443.

21. Tang K.-S. (2023). The characteristics of diagrams in scientific explanations: Multimodal integration of written and visual modes of representation in junior high school textbooks. Science Education, 107(3), 741-772.
22. Vaghari D.& Kabir E.& Henson R.N. (2022). Late combination shows that MEG adds to MRI in classifying MCI versus controls. NeuroImage, 252(nan), nan-nan.
23. Pasadas D.J.& Barzegar M.& Ribeiro A.L.& Ramos H.G. (2022). Locating and Imaging Fiber Breaks in CFRP Using Guided Wave Tomography and Eddy Current Testing. Sensors, 22(19), nan-nan.
24. Roheda S.& Krim H.& Riggan B.S. (2021). Robust Multi-Modal Sensor Fusion: An Adversarial Approach. IEEE Sensors Journal, 21(2), 1885-1896.
25. Planchuelo-G√≥mez √Å.& Garc√≠a-Azor√≠n D.& Guerrero √Å.L.& Aja-Fern√°ndez S.& Rodr√≠guez M.& de Luis-Garc√≠a R. (2021). Multimodal fusion analysis of structural connectivity and gray matter morphology in migraine. Human Brain Mapping, 42(4), 908-921.
26. Majji S.R.& Chalumuri A.& Kune R.& Manoj B.S. (2022). Quantum Processing in Fusion of SAR and Optical Images for Deep Learning: A Data-Centric Approach. IEEE Access, 10(nan), 73743-73757.
27. Sharma A.& Sharma K.& Kumar A. (2023). Real-time emotional health detection using fine-tuned transfer networks with multimodal fusion. Neural Computing and Applications, 35(31), 22935-22948.
28. Gil-Guevara O.& Bernal H.A.& Riveros A.J. (2022). Honey bees respond to multimodal stimuli following the principle of inverse effectiveness. Journal of Experimental Biology, 225(10), nan-nan.
29. Nooralishahi P.& Lopez F.& Maldague X.P.V. (2022). Drone-Enabled Multimodal Platform for Inspection of Industrial Components. IEEE Access, 10(nan), 41429-41443.
30. Mallol-Ragolta A.& Semertzidou A.& Pateraki M.& Schuller B. (2022). Outer Product-Based Fusion of Smartwatch Sensor Data for Human Activity Recognition. Frontiers in Computer Science, 4(nan), nan-nan.
31. Singh M.K.& Kumar S.& Bhatnagar G.& Saini D.& Ali M.& Sharma C.M.& Sharma N. (2022). A Blend of Analytical and Numerical Methods to Compute Orthogonal Image Moments over a Unit Disk. Wireless Communications and Mobile Computing, 2022(nan), nan-nan.
32. Barrett J.& Viana T. (2022). EMM-LC Fusion: Enhanced Multimodal Fusion for Lung Cancer Classification. AI (Switzerland), 3(3), 659-682.
33. Shah S.K.& Tariq Z.& Lee J.& Lee Y. (2021). Event-driven deep learning for edge intelligence (Edl-ei). Sensors, 21(18), nan-nan.
34. Prashantha S.J.& Prakash H.N. (2021). A Features Fusion Approach for Neonatal and Pediatrics Brain Tumor Image Analysis Using Genetic and Deep Learning Techniques. International journal of online and biomedical engineering, 17(11), 124-140.
35. Chirakkal S.& Bovolo F.& Misra A.R.& Bruzzone L.& Bhattacharya A. (2021). A General Framework for Change Detection Using Multimodal Remote Sensing Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14(nan), 10665-10680.
36. Kim G.& Moon S.& Choi J.-H. (2022). Deep Learning with Multimodal Integration for Predicting Recurrence in Patients with Non-Small Cell Lung Cancer. Sensors, 22(17), nan-nan.
37. Chakravarty A.& Misra S. (2021). Hydraulic fracture mapping using wavelet-based fusion of wave transmission and emission measurements. Journal of Natural Gas Science and Engineering, 96(nan), nan-nan.
38. Liu Q.& Kampffmeyer M.& Jenssen R.& Salberg A.-B. (2022). Multi-modal land cover mapping of remote sensing images using pyramid attention and gated fusion networks. International Journal of Remote Sensing, 43(9), 3509-3535.
39. Salama A.S.& Mokhtar M.A.& Tayel M.B.& Eldesouky E.& Ali A. (2021). A Triple-Channel Encrypted Hybrid Fusion Technique to Improve Security of Medical Images. Computers, Materials and Continua, 68(1), 431-446.
40. Dounis A.& Avramopoulos A.-N.& Kallergi M. (2023). Advanced Fuzzy Sets and Genetic Algorithm Optimizer for Mammographic Image Enhancement. Electronics (Switzerland), 12(15), nan-nan.
41. Baumann F.& Becker C.& Freigang V.& Alt V. (2022). Imaging, post-processing and navigation: Surgical applications in pelvic fracture treatment. Injury, 53(nan), S16-S22.

42. Geenjaar E.P.T.& Lewis N.L.& Fedorov A.& Wu L.& Ford J.M.& Preda A.& Plis S.M.& Calhoun V.D. (2023). Chromatic fusion: Generative multimodal neuroimaging data fusion provides multi-informed insights into schizophrenia. Human Brain Mapping, 44(17), 5828-5845.
43. Zhang T.& Ren J.& Li J.& Nguyen L.H.& Stoica P. (2022). RFI Mitigation for One-Bit UWB Radar Systems. IEEE Transactions on Aerospace and Electronic Systems, 58(2), 879-889.
44. Pearson H.C.& Wilbiks J.M.P. (2021). Effects of audiovisual memory cues on working memory recall. Vision (Switzerland), 5(1), nan-nan.
45. Juli√† i Juanola A.& Ruiz i Altisent M.& Boada i Oliveras I. (2022). An efficient and uniformly behaving streamline-based Œ°CT fibre tracking algorithm using volume-wise structure tensor and signal processing techniques. Computer Methods in Applied Mechanics and Engineering, 394(nan), nan-nan.
46. Ellen J.G.& Jacob E.& Nikolaou N.& Markuzon N. (2023). Autoencoder-based multimodal prediction of non-small cell lung cancer survival. Scientific Reports, 13(1), nan-nan.
47. Anilkumar P.& Venugopal P. (2023). An improved beluga whale optimizer,ÄîDerived Adaptive multi-channel DeepLabv3+ for semantic segmentation of aerial images. PLoS ONE, 18(10 October), nan-nan.
48. Srinivas P.V.V.S.& Mishra P. (2022). Human Emotion Recognition by Integrating Facial and Speech Features: An Implementation of Multimodal Framework using CNN. International Journal of Advanced Computer Science and Applications, 13(1), 592-603.
49. Schmitgen M.M.& Wolf N.D.& Sambataro F.& Hirjak D.& Kubera K.M.& Koenig J.& Wolf R.C. (2022). Aberrant intrinsic neural network strength in individuals with ‚Äúsmartphone addiction‚Äù: An MRI data fusion study. Brain and Behavior, 12(9), nan-nan.
50. Olsen A.S.& H√∏egh R.M.T.& Hinrich J.L.& Madsen K.H.& M√∅rup M. (2022). Combining electro- and magnetoencephalography data using directional archetypal analysis. Frontiers in Neuroscience, 16(nan), nan-nan.
51. Younus A.& Kelly A.& Lekgwara P. (2021). Minimally invasive extreme lateral lumbar interbody fusion (XLIF) to manage adjacent level disease ‚Äì A case series and literature review. Interdisciplinary Neurosurgery: Advanced Techniques and Case Management, 23(nan), nan-nan.
52. Tarigan D.G.P.& Isa S.M. (2021). A PSNR Review of ESTARFM Cloud Removal Method with Sentinel 2 and Landsat 8 Combination. International Journal of Advanced Computer Science and Applications, 12(9), 189-198.
53. Bello H.& Marin L.A.S.& Suh S.& Zhou B.& Lukowicz P. (2023). InMyFace: Inertial and mechanomyography-based sensor fusion for wearable facial activity recognition. Information Fusion, 99(nan), nan-nan.
54. Lai W.-S.& Shih Y.& Chu L.-C.& Wu X.& Tsai S.-F.& Krainin M.& Sun D.& Liang C.-K. (2022). Face deblurring using dual camera fusion on mobile phones. ACM Transactions on Graphics, 41(4), nan-nan.
55. Jeong B.& Lee J.& Kim H.& Gwak S.& Kim Y.K.& Yoo S.Y.& Lee D.& Choi J.-S. (2022). Multiple-Kernel Support Vector Machine for Predicting Internet Gaming Disorder Using Multimodal Fusion of PET, EEG, and Clinical Features. Frontiers in Neuroscience, 16(nan), nan-nan.
56. Pattanaik B.B.& Anitha K.& Rathore S.& Biswas P.& Sethy P.K.& Behera S.K. (2022). Brain tumor magnetic resonance images classification based machine learning paradigms. Wspolczesna Onkologia, 26(4), 268-274.
57. Asla N.& Kha I.U.& Albahussai T.I.& Almous N.F.& Alolaya M.O.& Almous S.A.& Alwheb M.E. (2022). MEDeep: A Deep Learning Based Model for Memotion Analysis. Mathematical Modelling of Engineering Problems, 9(2), 533-538.
58. Faragallah O.S.& Muhammed A.N.& Taha T.S.& Geweid G.G.N. (2021). Liver lesions and acute intracerebral hemorrhage detection using multimodal fusion. Intelligent Automation and Soft Computing, 30(1), 215-225.
59. Tam S.& Tanriover O.O. (2023). Multimodal Deep Learning Crime Prediction Using Tweets. IEEE Access, 11(nan), 93204-93214.
60. Akg√∞l ƒ∞. (2023). Mobile-DenseNet: Detection of building concrete surface cracks using a new fusion technique based on deep learning. Heliyon, 9(10), nan-nan.
61. Jensen D.M.& Zendrehrouh E.& Calhoun V.& Turner J.A. (2022). Cognitive Implications of Correlated Structural Network Changes in Schizophrenia. Frontiers in Integrative Neuroscience, 15(nan), nan-nan.

62. Sethanan K.& Pitakaso R.& Srichok T.& Khonjun S.& Weerayuth N.& Prasitpuriprecha C.& Preeprem T.& Jantama S.S.& Gonwirat S.& Enkvetchakul P.& Kaewta C.& Nanthasamroeng N. (2023). Computer-aided diagnosis using embedded ensemble deep learning for multiclass drug-resistant tuberculosis classification. Frontiers in Medicine, 10(nan), nan-nan.
63. Lawrance N.A.& Shiny Angel T.S. (2023). Image Fusion Based on NSCT and Sparse Representation for Remote Sensing Data. Computer Systems Science and Engineering, 46(3), 3439-3455.
64. Zahari Z.L.& Mustafa M.& Abdubrani R. (2022). The multimodal parameter enhancement of electroencephalogram signal for music application. IAES International Journal of Artificial Intelligence, 11(2), 414-422.
65. Abdelfatih B.& Ismail B.H. (2022). An Adaptive Image Fusion Algorithm in the NSST Based on CDF 9/7 for Neurodegenerative Diseases. Traitement du Signal, 39(4), 1379-1385.
66. Nakase K.& Takeshima Y.& Konishi K.& Matsuda R.& Tamura K.& Yamada S.& Nishimura F.& Nakagawa I.& Park Y.-S.& Nakase H. (2022). Usefulness of the Multimodal Fusion Image for Visualization of Deep Sylvian Veins. Neurologia Medico-Chirurgica, 62(10), 475-482.
67. Winterbottom T.& Xiao S.& McLean A.& Al Moubayed N. (2022). Bilinear pooling in video-QA: empirical challenges and motivational drift from neurological parallels. PeerJ Computer Science, 8(nan), nan-nan.
68. Wang Y.& Zeng D.& Wada S.& Kurihara S. (2023). VideoAdviser: Video Knowledge Distillation for Multimodal Transfer Learning. IEEE Access, 11(nan), 51229-51240.
69. Chuang C.-Y.& Lin Y.-T.& Liu C.-C.& Lee L.-E.& Chang H.-Y.& Liu A.-S.& Hung S.-H.& Fu L.-C. (2023). Multimodal Assessment of Schizophrenia Symptom Severity From Linguistic, Acoustic and Visual Cues. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31(nan), 3469-3479.
70. Jakobson Mo S.& Axelsson J.& Stiernman L.& Riklund K. (2022). Validation of dynamic [18F]FE-PE2I PET for estimation of relative regional cerebral blood flow: a comparison with [15O]H2O PET. EJNMMI Research, 12(1), nan-nan.
71. Iso-Mustaj√§rvi M.& Silvast T.& Heikka T.& Tervaniemi J.& Calixto R.& Linder P.H.& Dietz A. (2023). Trauma After Cochlear Implantation: The Accuracy of Micro-Computed Tomography and Cone-Beam Fusion Computed Tomography Compared With Histology in Human Temporal Bones. Otology and Neurotology, 44(4), 339-345.
72. Chugh A.J.S.& Patel M.& Chua L.& Arafah B.& Bambakidis N.C.& Ray A. (2021). Management of giant prolactinoma causing craniocervical instability: illustrative case. Journal of Neurosurgery: Case Lessons, 1(23), nan-nan.
73. Singh S.& Khosla A.& Kapoor R. (2023). Object tracking via a Novel Parametric Decisions based RGB-Thermal Fusion. International Journal of Image, Graphics and Signal Processing, 15(4), 1-18.
74. Meo C.& Franzese G.& Pezzato C.& Spahn M.& Lanillos P. (2023). Adaptation Through Prediction: Multisensory Active Inference Torque Control. IEEE Transactions on Cognitive and Developmental Systems, 15(1), 32-41.
75. Rathi S.& Kant Hiran K.& Sakhare S. (2023). Affective state prediction of E-learner using SS-ROA based deep LSTM. Array, 19(nan), nan-nan.