

Graph-Enhanced Multimodal Deep Learning for Disease Prediction: A Hybrid Approach

Ashwin M Nayak
Post Graduate Student
Dept. of MCA, DSCE
Bengaluru

ashwinnayak268@gmail.com

Dr. Vibha M B
Assistant Professor
Dept. of MCA, DSCE
Bengaluru

vibha-mcavtu@dayanandasagar.edu

Abstract—Disease prediction is crucial in healthcare, enabling early detection and proactive intervention for effective treatment and management. In this research, we explore a graph-enhanced multimodal deep learning approach along with a combination of Machine Learning and Natural Language Processing methods [9] for disease and symptom prediction using multimodal data. Two implementation approaches are investigated: multimodal deep learning and graph-based methods.

Through a thorough literature review, we establish a robust methodology that involves comprehensive data collection from diverse sources, including photos, text, and audio. Data preprocessing techniques, such as normalization, feature scaling, one-hot encoding, tokenization, and lemmatization, ensure compatibility and consistency across modalities.

Our study aims to enhance disease and symptom prediction accuracy through a hybrid approach that combines ML, NLP, and graph-enhanced multimodal deep learning. This approach assists healthcare professionals and individuals without specialized medical expertise in making informed decisions, ultimately improving patient outcomes. By integrating graph-based methods, the model achieves accurate disease prediction based on symptoms.

Keywords: Disease prediction, symptom prediction, machine learning, natural language processing, multimodal deep learning, graph-based methods.

I. INTRODUCTION

Combining Machine Learning (ML) with NLP methods [9] (Mewburn, 2018) has demonstrated significant potential in disease diagnosis and prediction. By analyzing textual data and extracting valuable insights, researchers have developed models that facilitate early disease detection and symptom prediction. As the field continues to evolve, novel approaches have emerged to further enhance prediction accuracy and improve patient outcomes.

Initially, researchers focused on ML and NLP approaches for text processing, successfully identifying trends and extracting useful elements for disease prediction from various sources, including medical records, clinical notes, and patient surveys. This early research laid a strong foundation for subsequent studies in the field.

Building upon the success of text processing approaches, researchers have now expanded their methodologies to

include multimodal deep learning and graph-based methods. By integrating information from different modalities, such as medical images, clinical notes, and patient-generated data, researchers aim to create a comprehensive representation that captures the complex relationships between symptoms and diseases.

In parallel, graph-based methods have gained attention in disease and symptom prediction. Graph Convolutional Networks (GCNs) or Graph Neural Networks (GNNs) are specifically designed to handle multimodal data represented as graphs. By representing entities as nodes and their relationships as edges, these methods effectively capture the dependencies and interactions between different modalities. This graph-based approach enables researchers to uncover hidden patterns and dependencies that may not be easily detected using traditional methods, leading to enhanced multimodal predictions and providing valuable insights for disease management.

II. LITERATURE REVIEW

In comparison to conventional ML algorithms, deep learning algorithms can provide predictions that are more sophisticated and complicated. To increase the precision of disease identification, they recommend investigating the use of deep learning algorithms.

In previous studies, a number of NLP techniques, such as symptom frequency analysis, similarity measurements, and clustering analysis, were used to improve disease identification accuracy.[6] (Akila, 2022)

To extract and process relevant details from textual data, such as symptoms, medical history, and patient descriptions, NLP techniques are used. These techniques give the framework the capacity to understand and evaluate the input data in a way that is comparable to how a person would interpret it. [6] (Akila, 2022)

The difficulties that the healthcare sector is facing because of the volume of unstructured, multimodal medical data that is growing. Traditional data warehouses are expensive, non-real-time, and have difficulty successfully integrating and exploring multimodal data.

Integrating EHRs into the disease identification process can improve the accuracy and completeness of the data. A patient's medical history, including past diagnoses, prescriptions, and treatments, can often be found in great detail in EHRs. [6]

An overview of multimodal machine learning as it is now in the context of precision medicine. The authors focused on the merging of diverse data to improve prediction and imitate clinical expert decision-making in their systematic analysis

of 128 studies published between 2011 and 2021. The paper emphasizes that the health industry has predominantly used single-modal data for machine learning while integrating multimodal data is a new area of research.[2] (Kline, 2022)

Neurology and oncology were determined to be the two medical specialties that use multimodal approaches the most frequently. In addition, the study highlights the drawbacks identified in the examined studies. Model fitting and generalizability problems are caused by small sample sizes.

The use of GNNs for multimodal causability. allowing the fusion of information and the definition of causal links between features using graph structures. The focus is on constructing a multimodal feature representation space to provide innovative interface methods and explanations.[4]

To address the task of predicting illness and their symptoms based on the advanced tools and technologies of Deep Learning, Machine Learning and NLP, we have come up with two distinct approaches namely Multimodal Deep Learning Approach and Graph-Based Methods approach.

2.1 Multimodal Deep Learning Approach:

- Process and extract features from the visual data using 2.1.1 Convolutional Neural Networks (CNNs).[7] (Laffitte, 2019)
- Process and extract features from the text and audio input Recurrent Neural Networks (RNNs).
- Use fusion techniques to merge the features that were derived from several modalities, such as concatenation or attention mechanisms.
- Create and train a multimodal deep learning model that uses the combined features to foretell illnesses and their symptoms.
- To enhance the performance of the model, take into account using relevant loss functions, such as cross-entropy or binary classification loss.

2.1.1 Convolutional Neural Networks:

By learning hierarchical features and patterns through convolutional layers, CNNs are designed to

interpret and analyze visual data, for instance medical images or scan reports.

The below mentioned formula of CNNs is used to draw features from image data

$$y = f(W * x + b)$$

Apply a convolution operation with a non-linear activation function (f), filter weights (W), input (x), and bias term (b) to draw features from image data.

2.1.2 Fusion Techniques:

Fusion techniques aim to create a unified representation of the multimodal data, which can provide a more comprehensive understanding of the underlying patterns and relationships. The below mentioned formula of fusion technique is used to depict fused feature vector

$$fused_features = [feature_1, feature_2, \dots, feature_n]$$

Concatenate the extracted features from different modalities to create a fused feature vector.

2.2 Graph-Based Methods Approach:

- Represent the multimodal data as a graph, with entities (e.g., patients, symptoms, diseases) as nodes and their relationships as edges.
- Utilize GCNs or GNNs to capture multimodal dependencies and interactions within the graph.
- Incorporate graph-based features and embeddings into the disease and symptom prediction task.
- Train the graph-based model using appropriate optimization techniques, such as graph convolution and backpropagation.
- Utilizing the two strategies outlined above, we can accurately predict disease using symptoms.

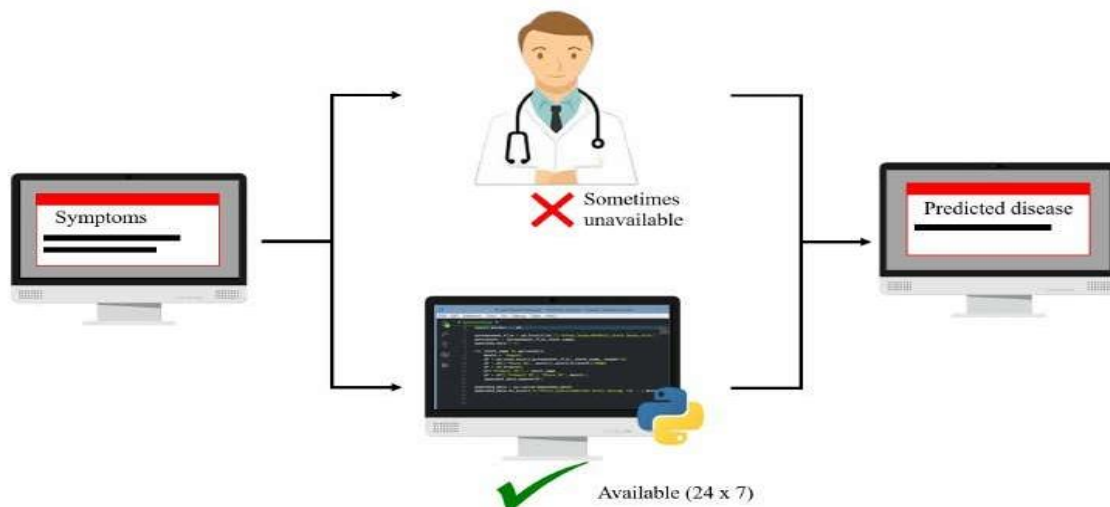


Fig. 1 The proposed approach for predicting disease. It's possible that the doctor won't always be available. But in the present world, this prediction mechanism is always available whenever it is needed. [1] (Keniya, 2020)

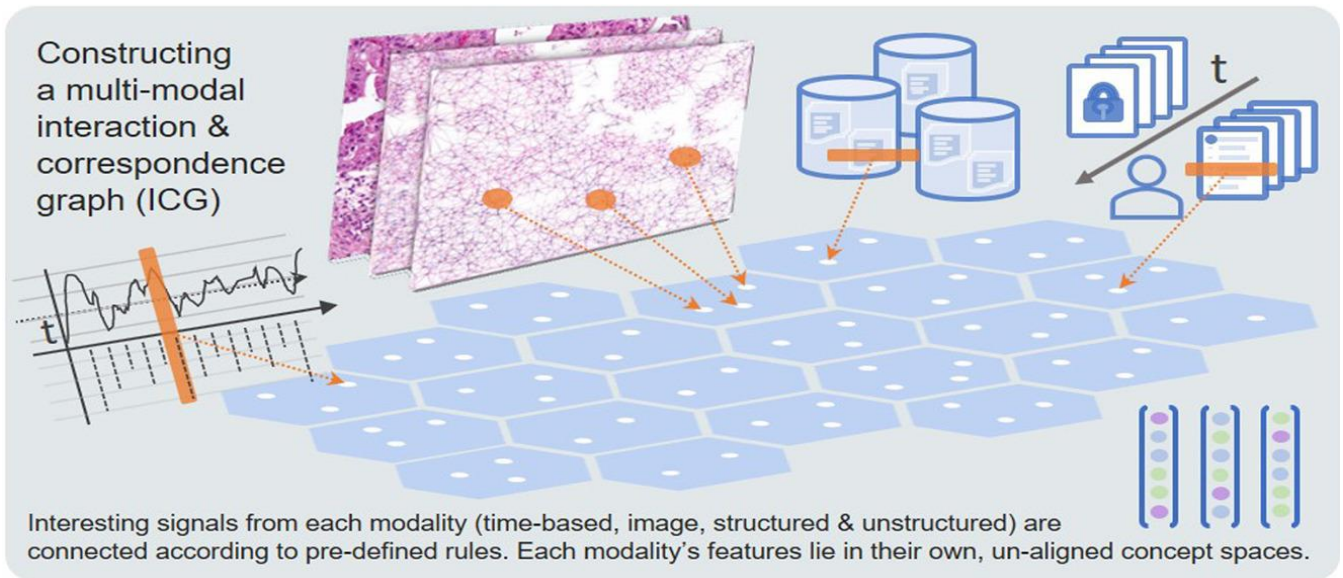


Fig. 2 Fusion of graphs. Interaction & Correspondence Graph (ICG) for Multimodal Data Integration [8] (Holzinger, 2021)

III. METHODOLOGY FOR PROPOSING THE IMPLEMENTATION OF HYBRID TECHNIQUE

The methodology employed in this study is grounded in an extensive literature review, forming the basis for developing a hybrid approach. By integrating insights from previous research, we have devised a methodology that combines a multimodal deep learning approach with a graph-based approach. The implementation steps for Graph-Enhanced Multimodal Deep Learning for Disease Prediction are as follows:

Step 3.1: Data Collection

Find a relevant dataset that has labels for the disease and symptoms together with multimodal data, such as photos, text, and audio. Ensure that the dataset covers a diverse range of diseases and symptoms to provide comprehensive training and evaluation.

Step 3.2: Data Preprocessing

- Preprocess and clean the multimodal data to ensure compatibility and consistency between modalities.
- Normalize and standardize the data to eliminate any biases or variation that might have an impact on the performance of the models.

3.2.1 Feature Scaling:

- Normalization:
 $x_{\text{normalized}} = (x - \text{mean}) / \text{standard deviation}$

Normalise the data by dividing by the standard deviation and removing the mean to standardise the characteristics

3.2.2 Feature Encoding:

- One-Hot Encoding: Create binary vectors from categorical variables in which each category is

represented by a single feature with a binary value (0 or 1).

- Label encoding: Assign distinct numerical labels to each category in a categorical variable.

3.2.3 Missing Data Handling:

- Find values that are missing and replace it with either the median, mean, or mode of available data for that feature.

3.2.4 Text Preprocessing:

- Tokenization: Split text into individual tokens (words, characters, or n-grams) for further processing.
- Stop word Removal: Remove commonly occurring words (e.g., "the," "is," "and") that may not carry significant meaning in the context.
- Lemmatization: Convert words to their base or dictionary form (lemma) to reduce inflectional variations.

3.2.5 Dimensionality Reduction:

- Principal Component Analysis: Create a fresh set of orthogonal features using the original features that captures the maximum variance in data. [11][13]
- t-SNE (t-Distributed Stochastic Neighbour Embedding): Reduce higher-dimensional data to a lower dimension using this technique.[13]

Step 3.3: Hybrid Model

Graph-Enhanced Multimodal Deep Learning

- Develop a hybrid architecture that integrates multimodal deep learning and graph-based methods to enhance disease and symptom prediction.
- Analyze and extract features from visual data using Convolutional Neural Networks as part of the multimodal deep learning strategy.
- Utilise RNNs to analyse and extract features from the text and audio input, following the multimodal deep learning technique.
- Apply fusion techniques to merge features extracted from multiple modalities, creating a fused feature vector for the hybrid model.
- Represent the multimodal data as a graph, with patients, symptoms, and diseases as nodes, and their relationships as edges, similar to the graph-based methods
- Capture multimodal dependencies and interactions within the graph using Graph Convolutional Networks or Graph Neural Networks, as discussed in the graph-based approaches section.
- Incorporate the graph-based features and embeddings into the disease and symptom prediction task within the hybrid model.
- Train the hybrid model using suitable optimization approaches like graph convolution and backpropagation to jointly learn multimodal and graph-based representations.
- To enhance the hybrid model's performance, consider relevant loss functions such as binary classification loss.[14]

Step 3.4: Evaluation Metrics

- Choosing a suitable evaluation metrics for evaluating the performance of the disease and symptom prediction models in the hybrid framework. [11]
- Examine the model's performance using proposed evaluation metrics and the test data for evaluating the model's performance.

IV. CONCLUSION

The study emphasizes the significance of leveraging multimodal data and advanced techniques in healthcare information systems, offering valuable insights for future research and development in this field.

The successful utilization of ML and NLP [9] (Mewburn, 2018) approaches for disease and symptom prediction using multimodal data. The work primarily used graph-based techniques and multimodal deep learning.

The hybrid approach synergistically combines multimodal deep learning techniques, including CNNs and RNNs, with the power of graph-based methods, specifically GCNs or GNNs. This integration aims to harness the wealth of information present in diverse modalities such as visual, textual, and auditory data while capturing intricate dependencies and interactions through graph-

based representations. By fusing features extracted from different modalities, our hybrid model achieves a comprehensive representation of input data, facilitating a more holistic understanding of patients' health conditions. The incorporation of graph-based features and embeddings enables the model to effectively discern relationships and dependencies among patients, symptoms, and diseases, thus augmenting its predictive capabilities.

While the proposed hybrid approach is still in the proposal stage and awaits implementation and evaluation, we firmly believe it will outperform individual multimodal deep learning or graph-based techniques. The hybrid model will be rigorously evaluated using standard metrics to ensure a fair and comprehensive analysis.

V. REFERENCES

- [1] Keniya, R., Khakharia, A., Shah, V., Gada, V., Manjalkar, R., Thaker, T. H., Warang, M., & Mehendale, N. (2020). Disease Prediction From Various Symptoms Using Machine Learning. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3661426> .
- [2] Kline, A., Wang, H., Li, Y. et al. Multimodal machine learning in precision health: A scoping review. *npj Digit. Med.* 5, 171 (2022). <https://doi.org/10.1038/s41746-022-00712-8> .
- [3] Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., Yang, Y., & Zhao, Y. (2022, September). Multi-Modal Graph Learning for Disease Prediction. *IEEE Transactions on Medical Imaging*, 41(9), 2207–2216. <https://doi.org/10.1109/tmi.2022.3159264> .
- [4] Andreas Holzinger, Bernd Malle, Anna Saranti, Bastian Pfeifer, Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI, *Information Fusion*, Volume 71, 2021, Pages 28-37, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2021.01.008> .
- [5] Zhang, Y., Sheng, M., Liu, X., Wang, R., Lin, W., Ren, P., Wang, X., Zhao, E., & Song, W. (2022, August 26). A heterogeneous multi-modal medical data fusion framework supporting hybrid data exploration. *Health Information Science and Systems*, 10(1). <https://doi.org/10.1007/s13755-022-00183-x>
- [6] Akila, S. (2022). Disease Identification using Machine Learning and NLP. *Journal of Science Technology and Research (JSTAR)* 3 (1):78-92.
- [7] Laffitte, P., Wang, Y., Sodayer, D., & Girin, L. (2019, March). Assessing the performances of different neural network architectures for the detection of screams and shouts in public transportation. *Expert Systems With Applications*, 117, 29–41. <https://doi.org/10.1016/j.eswa.2018.08.052> .

- [8] Holzinger, A., Malle, B., Saranti, A., & Pfeifer, B. (2021, July). Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI. *Information Fusion*, 71, 28–37. <https://doi.org/10.1016/j.inffus.2021.01.008> .
- [9] Mewburn, I., Grant, W. J., Suominen, H., & Kizimchuk, S. (2018, July 4). A Machine Learning Analysis of the Non-academic Employment Opportunities for Ph.D. Graduates in Australia - Higher Education Policy. SpringerLink. <https://doi.org/10.1057/s41307-018-0098-4> .
- [10] Sameer, M., . O., Gupta, R., Tyagi, R., & Mishra, A. (2022, April 30). Common Ailments Possibility Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 10(4),1897–1902. <https://doi.org/10.22214/ijraset.2022.41588> .
- [11] Diamantaras, K., Duch, W., & Iliadis, L. S. (Eds.). (2010). *Artificial Neural Networks – ICANN 2010. Lecture Notes in Computer science*. <https://doi.org/10.1007/978-3-642-15819-3> .
- [12] Choi, Jung-gu & Ko, Inhwan & Han, Sanghoon. (2021). Depression Level Classification Using Machine Learning Classifiers Based on Actigraphy Data. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2021.3105393 .
- [13] Bornet, A., Proios, D., Yazdani, A., Jaume-Santero, F., Haller, G., Choi, E., & Teodoro, D. (2023, June 5). Comparing neural language models for medical concept representation and patient trajectory prediction. medRxiv. <https://doi.org/10.1101/2023.06.01.23290824> .
- [14] *Database Systems for Advanced Applications*. (n.d.). SpringerLink. <https://link.springer.com/book/10.1007/978-3-031-00123-9> .