RESEARCH PAPER

Semantic Similarity Learning for Medical Literature: A Multimodal Framework Combining LLM and Deep Metric Learning

Deepika A[1*], Rajeswari M[2]

[1*] Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women, India
deepuarumugam22@gmail.com
[2] Assistant Professor, Department of Computer Science, PSGR Krishnammal College for Women, India

Abstract
There is an increasing need for artificial intelligence in the field of medical related learning frameworks which is capable of efficiently handling the multimodal data which is under resource constrained situations. Conventional machine learning and deep learning algorithms are usually depending on the large labelled dataset and limited environment with less applicability for real world applications. The adoption of the Artificial intelligence in the field of healthcare is the need for the learning and effective handling of the multimodal healthcare data. Traditional machine learning and deep learning algorithms depends on the large-scale datasets with label, limiting the usage of the current clinical environment. This will limit the applicability of the real-world clinical environments by the data lackness and variance in the domain. In this work, Large Language Models are used with deep metric learning framework for multimodal medical journal classification. In this work, Large Language Model is implemented using deep metric learning framework for the multimodal dataset. The multimodal dataset comprises of text, table and images from online repository. The data manually downloaded from the PubMed website is used in this work. The proposed LLM guide framework outperformed all other models trained by various benchmark datasets.

Keywords- Cross-Modal Embeddings, Deep Metric Learning, Explainable AI in Healthcare, Few-Shot Medical Classification, Large Language Models, Multimodal Learning
How to cite this article: Deepika A, Rajeswari M. Semantic Similarity Learning for Medical Literature: A Multimodal Framework Combining LLM and Deep Metric Learning. Int J Drug Deliv Technol. 2026;16(13s): 810-814. DOI: 10.25258/ijddt.16.13s.88.

1. Introduction

The transformation of the publication of the research articles in digital repositories led to the rapid accumulation of online data which includes diverse medical data which includes medical images, clinical reports, digital health records. Nowadays handling the multimodal data has become tedious to increase the diagnostic accuracy, taking clinical decisions and the outcome of the patient. Recent advances in deep learning algorithms makes significant progress in analyzing the medical data. Still the existing approaches highly depends on structured with labelled datasets and existing task specific models. These architectures gained less accuracy on handling real life clinical environments especially less resource with data scarcity. [1]

The traditional learning algorithms adopt to the closed set classification paradigm the depends on softmax based supervision. These algorithms are effective in controlled settings but generate less generalization when it is used in unknown or new classes, changing from one domain to another domain, or highly imbalanced datasets. In healthcare applications, these limitations are generally problematic because the medical knowledge expands nowadays with the clinical data becomes diverse.[2] The cost required to handle the high-quality medical data is too high nowadays as reduces the scalability of the machine learning and deep learning algorithms.

Deep Metric Learning model is one of the most useful algorithms for which is used to learn the similarity in the datasets instead of splitting them from the fixed classes. This model places the related data items together in the special feature space. This pattern of the Deep metric algorithm is used to handle the medical data where it is used to analyse the newly generated cases with the old clinical records for the reference. Anyhow the model is used mostly for only one type of data with understanding of the simple representations. [3] Latest models like Large Language Models (LLMs) are used in this Natural Language Processing domain to handle various text related data such as biomedical literatures and clinical records. The advent of this LLMs has become significant as this is pretrained on large scale corpora and capable of using few-shot generalization which provides high contextual embedding of the given datasets. This model is capable of understanding the domain specific knowledge and semantic meaning of the given data. These characteristics of the LLM algorithms is suitable for the medical text data analysis in which the interpretation of the data is essential. This model is used for the classification, question answering tasks, sentimental analysis tasks. [4]

By analyzing the working of LLM and Deep Metric Learning algorithms, in this work LLM guided metric learning framework is implemented for multimodal medical journal classification. This is used to handle the challenges imposed by the limited labelled data, domain shifts and interpretation of the data. The proposed work combines the Large Language Models based text embeddings with Convolutional Neural Network with derived visual representations into unified metric space

*Author for Correspondence:deepuarumugam22@gmail.com

with optimized metric space which is optimized by metric learning. [5]

Concentrating on meaning-based similarity than the rigid class discrimination, this work is efficient in handling different types of data and in different environments. The few-shot learning with the adaptation to the different domains makes the model to efficiently adapt to the various clinical situations with less super vision. [6-9]

Multiple experimental evaluations were conducted on many benchmark datasets to analyse the efficiency of the proposed model under few-shot learning and cross domain state. This work is capable of handling data efficiently and provide effective solution for the multimodal journal classification by increasing the real-world medical applications.

2. Related works

This section discusses about the related works in this domain. Amita Kumari, et.al., analysed the ability of the various GPT models such as chatgpt, Google Bard, Microsoft Bing. The GPT models are used to provide the solution for the haematology case related questions of the curated dataset from 50 scenarios. From the implemented works the chatgpt outperforms all other GPT models.

Lidia Moura et.al., proposed a work by implementing Large Language models to gather the information about neurology related treatments. This GPT model analyse the cost for the treatment, patient care, and safety of the patients. Ehsan Ullah et. al., analysed the barriers and challenges created by using the latest AI models for providing the solution to the medical related cases. This review paper identified various impact of the using the latest innovation on the medical scenario and concerns about the privacy of the patient's data.

Abdulrahmaan Tamsah et. al., presented this work by revealing the challenges in adopting to the latest advancements in artificial intelligence algorithms in healthcare field. Also provided the broad understanding of the latest model DeepSeek usage on healthcare. Bertalan Mesko proposed a work to examine the impact of the multimodal LLMs in the healthcare field and scrutinized the future impact of implementing the AI algorithm on medical field.

Zohar Elyoseph et.al., proposed a work by comparing the various perspective of the generative AI and mental health cases. This work deeply analysed the Schizophrenia, the chronic illness disease using the advanced algorithms. By using this algorithm, they tried create solution for the disease as this has severe symptoms.

Zohar Elyosephet.al., implemented multimodal data like visual and textual data to capture the emotions of humans by using the generative AI. This work examines and interpreted human emotions which includes beliefs and intentions. Reading the mind in the eyes test created by Baron – Cohen and their colleagues to create emotional awareness scale used to analyse the efficiency of the Large Language Models.

3. Methodology

3.1 Datasets

To analyse the effectiveness of the model proposed in this work, this framework is applied on various datasets available on medical field. The datasets used are diverse in nature which comprises of multiple modals.

The dataset used to train the algorithm are benchmark datasets that are used widely in all medical related research work. Megsegbench is the dataset used to train the algorithm. This is one of the famous datasets used to evaluate the efficiency of the algorithm. This dataset comprises of CT scan, MRI scan images, endoscopy data, X-ray data. The images are preprocessed using the normalization to make the dataset to fit into the visual encoders used in this work. The trained model is implemented in the PubMed dataset downloaded manually downloaded from the PubMed website. 2500 medical abstracts downloaded by using the keyword Cancer and its types. The images and tables are downloaded from the same website from cancer related research articles published in PubMed website. [8,12]

3.2 Deep Metric Learning Fundamentals

The concepts of the proposed algorithms deep metric learning and large language models and multimodal representation are explained in this section. The methodology foundation is explained in following sections.

3.2.1 Metric Spaces

Deep Metric Learning algorithm aims to understand the embedding function that maps the data into the sequential metric space where the space between them is the similarity of the data. A distance function used is Euclidean distance which is used to calculate the similarity of the given dataset. Deep Metric Learning (DML) aims to learn an embedding function that maps input samples into a continuous metric space where distances reflect semantic similarity. [13]

The main goal of the DML is to similar samples are with the less distance in the embedding space and the non-similar data plotted with larger distance. The traditional classification approaches that depend on the fixed boundaries that was given in the algorithm while training but this metric learning algorithm splits the dataset according to the sematic relationship between the dataset. This enables the increased generalization which involves imbalances in the data, previously unknown categories, and less labelled data, and this property makes the proposed algorithm suitable for the medical related datasets.[14]

3.2.2 Contrastive and Triplet Loss

The embedded space is optimized by using the loss function with DML frameworks that can directly impose the similarity constraints. The contrastive pair of the model which produce pair of inputs. They are positive input and negative pair. This positive pair that represents the similar items and the negative pair that represents the different classes.

The loss function

$L = (1-Y).D^2 + Y.\max(0,m-D)^2$

In the equation, D is the distance between two embeddings, Y represents the similar and dissimilar pair according to the values. If the value is 0 that means they are similar pair, if the value is 1 they represent the dissimilar pair. The value m represents the margin value that is the distance between the different classes. [15]

The triplet loss is also called as the relative comparison learning. In this, the model compares three samples at a time instead of simply checking the two values that is similar or different values. This loss contains three values Anchor(A), Positive (P), Negative(N). The anchor is the reference sample, the positive is the same class as the anchor, negative is the different class other than the anchor.

$L = \max(0, D(A, P) – D(A, N) + m)$

Where D(A,P) represents the distance between Positive and Anchor pair, D(A, N) represents the distance between Negative and Anchor pair and m represents margin. To reduce the loss the function should be the anchor and positive pair distance is smaller than the anchor and negative pair distance. Triplet loss is also used because of its capability of finding the similarity relationships and generate the large discriminative embeddings. [16]

3.3 Large Language Models as Semantic Encoders

3.3.1 Embedding Properties

Large Language Models (LLMs) are already trained with the large text corpora by the use of self-supervised objectives which makes the model to find high syntactic and the semantic structures. In this work, the LLM is used as the encoders. So, they convert the input text into vector representations that encode the contextual meaning, with the domain specific understanding, and long-range dependencies. In this work, the text input is analysed by our model and produces an embedding that is the semantic representation for the downstream learning task. The embedding that has strong understanding across different models and across different domains, and different types of data. This understanding makes the algorithm perform effective medical text analysis without label. In this framework, the embedding generated by the LLM which preserve the semantic similarity like the meanings of the text is analysed and the they are mapped accordingly. This property groups the data naturally with the objectives of the learning.

3.3.2 Few-Shot Capabilities

The few-shot learning is the concept of teaching the algorithm to perform new task by providing a smaller number of training data instead of providing large corpora. These few-shot learning is used for medical text dataset, as training the large dataset is quite expensive, and time consuming to understanding the context of the medical terms.

The special feature of the LLMs is their capability of effectively performing in low resource environments. LLMs pre training with the diverse dataset and it gain the general knowledge about the text and domain.

Through pre-training on diverse corpora, LLMs acquire generalized linguistic and domain knowledge that can be leveraged with minimal task-specific supervision. In the context of metric learning, this enables the construction of semantically meaningful embeddings even when only a small number of labelled examples are available. Consequently, LLMs serve as powerful semantic anchors that facilitate robust representation learning in data-scarce medical environments.

4. Proposed Methodology

This section explains about the proposed framework of this work LLM guided deep metric learning framework for multimodal medical journal classification. This is proposed to address the challenges posed by the related to classifying the unlabelled multimodal dataset.

4.1 Overall Framework Architecture

The methodology consists of three components a text encoder, a visual encoder, and the metric embedding space. For the text embedding LLM was implemented, CNN is used for the image embedding and these both embedding is projected on metric space. The metric space where the data projected was shared by both text and image embeddings.

4.1.1 LLM for Text Embedding

Text embedding is the process of converting the given text data into numerical values. The numerical vectors are feed into the algorithms as input so that the algorithm can understands the depth meaning of the data and the relationships between each and every word in the dataset. In this work the dataset is downloaded from the PubMed website. The research articles with abstract, images and tables from the research articles are manually downloaded from the PubMed website. These inputs are clinically encoded using the pre trained Large Language Models which maps the input data into semantic embedding. The algorithm captures the semantic meaning of the dataset, the feature extraction on low – resource environments.

4.1.2 CNN for Image Embedding

Image embedding is the concept of converting the given image dataset into a numerical vector of fixed length which represents the image content. The CNN algorithm is one of the important algorithms for handling image data. This algorithm automatically analyses the patterns in the dataset such as edges, shapes, textures and objects. ImageNet is used to enhance the generalization.

4.1.3 Shared Metric Space

The embeddings used in this work are projected into the metric space. This is shared by both text and image embeddings. In this the semantic similarities across modalities are measured by the same distance function. This space enables the cross-modal comparison, retrieval and the classification of the dataset according to the distance calculated by the distance function.

4.1.4 Evaluation Metrics

The implemented model is evaluated using the standard metrics that are commonly used in biomedical research. Following are the evaluation metrics used. Accuracy, F1 Score, AUC curve are used to evaluate the algorithms. The F1 score which checks for the balances between precision, recall of the datasets. AUC curve which is

used to assess the capability of the model ability to discriminate the classes across the various threshold settings.

5. Results and Analysis

The comprehensive results of the proposed model is discussed in this section. The LLM – guided deep metric algorithm outperformed all other algorithms compared. This work mainly focused on overall analysis of classification performed by the models implemented and the general working of the few - shot learning on medical image dataset and multimodal dataset. The table below shows the accuracy of the various algorithms compared in this work. Table 1, Table 2, Table 3 depicts the accuracies of various algorithm implemented in this work.

Table 1. Accuracies of the various algorithms.

Model Type	Model	Learning Strategy	Accuracy (%)
CNN	ResNet-50	Softmax (baseline)	78.4
CNN	EfficientNet-B0	Softmax (baseline)	80.1
Transformer-based	BERT	Softmax (text-only)	81.6
Multimodal	CNN + BERT	Late Fusion	83.2
Metric Learning	CNN + Text Encoder	Triplet Loss	85.7
Proposed	CNN + LLM	Deep Metric Learning	89.3

Table 2. Few-Shot Learning Accuracy (%)

Shots per Class	CNN Baseline	Transformer Baseline	Proposed LLM + DML
5-shot	62.8	65.4	74.6
10-shot	68.9	71.2	81.3
20-shot	74.5	76.8	85.9

Table 3. Cross-Domain Generalization Accuracy (%)

Model	Accuracy (%)
CNN (Softmax)	70.6
Multimodal Transformer	74.8
Metric Learning (no LLM)	80.2
Proposed LLM + DML	86.7

5.1 Quantitative Performance Comparison

The proposed model in this study outperforms the other algorithms implemented like CNN based and transformer-based algorithms. Firstly, the model was trained using the traditional softmax model. This model shows the promising accuracy when it is applied on the benchmark dataset but in the multimodal dataset it shows lesser accuracy. The proposed algorithm produced stable performance on the same dataset with good accuracy, F1 score, and AUC.

In few-shot learning methods, the model other baseline models as the baseline model experienced overfitting. The pretrained embedding using LLM generates semantic pattern in the given dataset and the outcomes of these models shows the suitability of the proposed models suitability for limited resource environments. Apart from the classification, the proposed model efficient retrieval of the semantics in the multimodal data such as medical images and clinical text data.

6. Conclusion

This proposed framework implemented on multimodal medical journal classification addresses the lack of efficient algorithms in the field of less resource environments. By combining the visual embedding with the text embedding the proposed work shows the high accuracy on the multimodal benchmark datasets. Various experiment results indicate the best performance of the few-shot learning and cross domain environments. This work also analyses the Large Language model based deep metric learning is performing effectively on the medical healthcare datasets. In future the LLM can be implemented for various modalities and various domains to analyse further strong understanding of diverse datasets.

References

1. AlSaad R, Abd-Alrazaq A, Boughorbel S, Ahmed A, Renault MA, Damseh R, Sheikh J, “Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook”. J Med Internet Res. 2024 Sep 25;26:e59505. doi: 10.2196/59505. PMID: 39321458; PMCID: PMC11464944.
2. Meskó B, “The Impact of Multimodal Large Language Models on Health Care's Future” J Med Internet Res. 2023 Nov 2;25:e52865. doi: 10.2196/52865. PMID: 37917126; PMCID: PMC10654899.

3. Ding Z, Wei R, Xia J, Mu Y, Wang J, Lin Y. "Exploring the potential of large language model-based chatbots in challenges of ribosome profiling data analysis: a review." Brief Bioinform. 2024 Nov 22;26(1):bbae641. doi: 10.1093/bib/bbae641. PMID: 39668339; PMCID: PMC11638007.
4. Limbu MS, Xiong T, Wang S. "A review of Ribosome profiling and tools used in Ribo-seq data analysis." Comput Struct Biotechnol J. 2024 Apr 22;23:1912-1918. doi: 10.1016/j.csbj.2024.04.051. PMID: 38721586; PMCID: PMC11076270.
5. Preiksaitis C, Ashenburg N, Bunney G, Chu A, Kabeer R, Riley F, Ribeira R, Rose C. "The Role of Large Language Models in Transforming Emergency Medicine: Scoping Review." JMIR Med Inform. 2024 May 10;12:e53787. doi: 10.2196/53787. PMID: 38728687; PMCID: PMC11127144.
6. King DR, Nanda G, Stoddard J, Dempsey A, Hergert S, Shore JH, Torous J. "An Introduction to Generative Artificial Intelligence in Mental Health Care: Considerations and Guidance." Curr Psychiatry Rep. 2023 Dec;25(12):839-846. doi: 10.1007/s11920-023-01477-x. Epub 2023 Nov 30. PMID: 38032442.
7. Claman D, Sezgin E." Artificial Intelligence in Dental Education: Opportunities and Challenges of Large Language Models and Multimodal Foundation Models." JMIR Med Educ. 2024 Sep 27;10:e52346. doi: 10.2196/52346. PMID: 39331527; PMCID: PMC11451510.
8. A, Deepika and N, Radha, Performance Analysis of Abstract based Classification of Medical Journals using Ensemble Methods (May 25, 2021). Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021), Available at SSRN: https://ssrn.com/abstract=3852605 or http://dx.doi.org/10.2139/ssrn.3852605
9. Choudhury A, Chaudhry Z. "Large Language Models and User Trust: Consequence of Self-Referential Learning Loop and the Deskilling of Health Care Professionals." J Med Internet Res. 2024 Apr 25;26:e56764. doi: 10.2196/56764. PMID: 38662419; PMCID: PMC11082730.
10. KAYA, M.; BİLGE, H.Ş. "Deep Metric Learning: A Survey." Symmetry **2019**, 11, 1066. https://doi.org/10.3390/sym11091066
11. X. Jiang et al., "Deep Metric Learning Based on Meta-Mining Strategy With Semiglobal Information," in IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 4, pp. 5103-5116, April 2024, doi: 10.1109/TNNLS.2022.3202571.
12. Deepika, A., Radha, N. (2022). "Performance Analysis of Abstract-Based Classification of Medical Journals Using Machine Learning Techniques." In: Smys, S., Bestak, R., Palanisamy, R., Kotuliak, I. (eds) Computer Networks and Inventive Communication Technologies . Lecture Notes on Data Engineering and Communications Technologies, vol 75. Springer, Singapore. https://doi.org/10.1007/978-981-16-3728-5_47
13. Kumari A, Kumari A, Singh A, Singh SK, Juhi A, Dhanvijay AKD, Pinjar MJ, Mondal H. "Large Language Models in Hematology Case Solving: A Comparative Study of ChatGPT-3.5, Google Bard, and Microsoft Bing." Cureus. 2023 Aug 21;15(8):e43861. doi: 10.7759/cureus.43861. PMID: 37736448; PMCID: PMC10511207.
14. Jitha P. Nair. (2024). "Incorporating Seasonal Trends for River Water Quality Prediction Models Using Deep Learning Algorithms." International Journal of Intelligent Systems and Applications in Engineering, 12(21s), 3987.
15. Moura L, Jones DT, Sheikh IS, Murphy S, Kalfin M, Kummer BR, Weathers AL, Grinspan ZM, Silsbee HM, Jones LK Jr,, Patel AD. "Implications of Large Language Models for Quality and Efficiency of Neurologic Care: Emerging Issues in Neurology." Neurology. 2024 Jun 11;102(11):e209497. doi: 10.1212/WNL.0000000000209497. Epub 2024 May 17. PMID: 38759131.
16. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos EG. "Evaluation of the Performance of Generative AI Large Language Models ChatGPT, Google Bard, and Microsoft Bing Chat in Supporting Evidence-Based Dentistry: Comparative Mixed Methods Study." J Med Internet Res. 2023 Dec 28;25:e51580. doi: 10.2196/51580. PMID: 38009003; PMCID: PMC10784979.