Leveraging Artificial Intelligence for Multi-Modal Learning in Healthcare Applications

*Yogesh Tarachand Patil[1], Pallavi Soni[2], Vivek Verma[3], Deepak Bishnoi[4]

Department of Computer Science and Application, Vivekananda Global University, Jaipur, India[1,2,3,4]
yogi007orama@gmail.com[1], anshom764@gmail.com[2], vv150077@gmail.com[3], deepbishnoi87@gmail.com[4]

Abstract.
With the aid of Artificial Intelligence (AI), information technology has enhanced the illness diagnosis and prognosis in the overall healthcare processes by accelerating clinical decision-making. This research work presents a new concept, a multi-modal learning approach, for the diagnosis of important medical conditions like lung diseases, diabetes, heart diseases, and kidney diseases. This tweaking method produces a hybrid model by incorporating the potential of XGBoost for prediction with a mixture of logistic regression for understanding. This approach of analysis enables a comprehensive analysis of numerous patient factors incorporated in a clinical case. The proposed ensemble model is superior to the individual models and proves its ability to identify various diseases with an accuracy of 94%. These findings show that ensemble learning eradicates the chances of enhancing diagnostic and prognostic classification accuracy in real-world practical areas, and it fills the existing gap of escalating the precision of diagnosing tools prevalent in the healthcare industry, thereby laying a strong foundation for the application of multi-stream intelligent systems

Keywords: Multi-Modal, Prediction, Ensemble, Disease, Model Performance, Ensemble.

N. Vyas et al. (eds.), *Proceedings of the International Conference on Advancements in Computing Technologies and Artificial Intelligence (COMPUTATIA 2025)*, Advances in Intelligent Systems Research 189,
https://doi.org/10.2991/978-94-6463-700-7_25

1 Introduction

An improvement in Artificial Intelligence (AI) has impacted several sectors, including the healthcare sector [1]. The ability to improve healthcare diagnosis and decision-making has expanded significantly over the last few years due to the availability of multimodal data from sources including electronic medical records, imaging, genetics, and wearable devices [2]. This work aims towards applying AI in multi-modal learning for the health care domain that includes Renal problems, Heart diseases, Diabetes, and Lung diseases using ensemble models of Logistic regression and XGBoost[3]. Many papers also demonstrated that AI can enhance the patients' situation and increase diagnosis precision, so its integration into healthcare becomes a rather promising area [4]. XGBoost, an ensemble learning technique, has often been used for its capability to minimize the error rate when a model has non-linear data and to give high accuracy, and logistic regression, a statistical model [5], is preferable for its good interpretability and binary classification [6]. While the integration of models that leverage their respective strengths has been the focus of recent analysis, ensemble learning methodologies that enhance diagnostic capabilities have been developed [7]. In this study, to propose a new approach identifying four significant medical diseases involving kidney diseases, heart diseases, diabetes, and lung diseases, we have adopted the ensemble model of logistic regression with XGBoost [8]. The increased prevalence of long-term diseases and the need to minimize mortality rates and improve patients' well-being due to the accuracy of diagnosis are motivating factors behind this research [9]. Usually, current diagnostic technologies fail at quickly processing and interpreting multi-modal data, which often results in a delayed or incorrect diagnosis [10].

2 Literature Review

This shift to assume explanatory power by submitting additional, richer kinds of data to analytic methods is an emergent feature of how artificial intelligence (AI) integration in health care has brought clinical utility and forecast precision into sharper focus. Another interesting aspect of the HAIM approach is that it has a generic format for the input, allowing it to work with text, image, and time series data, as well as tabular data. Among the 14,324 model types tested by the authors [11], overall performance was determined for 12 prediction tasks on 11 data sources using the HAIM-MIMIC-MM dataset (34,537 samples). Eight studies with overall performance improvement of 6-33% compared to the single-source approach were identified; chest pathology diagnosis, predicting length of stay, and 48-hour mortality were important findings. By means of Shapley values, modality-specific contributions have been made, revealing unique data relevance for different tasks.

However, despite such a great potential of the HAIM framework in clinical applications, it has some shortcomings that can prevent it from being used in resource-limited settings, such as dependence on large multimodal datasets and computational requirements. However, the claimed strength of the approach is flexibility and prediction accuracy, which can present a basis for the future development of multimodal systems in healthcare AI.The ability of traditional machine learning methods used in healthcare to mimic an accurate clinician's approach of integrating multiple information sources for decision-making has only been limited by the use of single-modality data. This is, of course, a limitation since a model trained on a single type of input cannot take into account all these functionalities at once, and this is where multimodal machine learning excels since it enables the usage of input sources such as text, time series, imagery, and tables all at once to create much more comprehensive models. Creation of a model is a multistage process, and it is now easier to perform the pre-training, fine-tuning, and assessment phases.

Three forms of data fusion are early, middle, and late fusion, which each have their advantages and, at the same time, their limitations in aspects of computational intensity and performance gains. After seven pilot studies and 17 multimodal clinical datasets were analyzed by the author [12], several data modalities for specific predicting tasks were identified. The integration of imaging and tabular data turned out to be the primary concern of slightly more than 50 research studies analysed. While loosely coupled multimodal approaches substantially enhance contextual relevance and prediction precision, their performance remains task- and data-dependent. Difficulties, including data harmonization, processing overhead, and modality-specific limitations, are some of the challenges, reinforcing the need for focused strategies for healthcare applications. Previous existing work done related to our work is shown in Table 1.

Table 1 Existing work done

Sr No.	Year of Publication	Proposed Work	Techniques Used	Limitations
[13]	2023	Created a new deep learning model ResNet50 for multi-modal medical image classification. Model trained on 28,378 images from the dataset and got an accuracy of 98.61%.	Deep Learning (ResNet50), CNNs	Concentrates on image classification only and does not consider multi-modal data fusion, yet, it explores the limitations of deep learning in clinical practice.

[14]	2023	Developed an anomaly detection system in healthcare using auto encoders while perspectively focusing on interpretability and ethical concerns. Protects data through ethical concerns of data privacy.	Deep Learning (Auto encoders), Feature Importance Analysis	Prone to anomaly detection only, consideration of multi-modal data integration, and potential modes of implementing and deploying the solution in different contexts of healthcare.
[15]	2021	Suggested an MMRL algorithm to enhance the energy efficiency of IoT based healthcare devices. Centered on the resilience of compression, low energy, and minimal delay.	IoT, Edge Computing, Multi-Modal Reinforcement Learning (MMRL)	Focused mainly on reducing energy consumption within IoT devices.

3 Methodology

The methodology offers a theoretical concept about applying AI to achieve multi-modal learning in healthcare to improve diagnostic accuracy and clinical decision support.

3.1 Data Collection

This work started with gathering multi-modal healthcare data from trustworthy, open-source data sources, including genomic databases, wearable sensor datasets, medical imaging repositories, and EHRs. To ensure the external validity of the suggested model, data included disease types such as kidney, heart, diabetes, and lung diseases. To be specific, there is a set of structured data such as numerical laboratory results combined with unstructured data like medical photos and word documents for analysis. Data quality, data constancy, and ethical information structures such as HIPAA and GDPR remain the basic foundations in the formation of proper AI. Increased data size and dimensionality improve operational generalisation, and the elimination of specific institutions minimises bias. The data collected is authentic and well labeled.

3.2 Data Preprocessing

The considerations associated with the heterogeneity of multi-modal healthcare data were addressed by applying the most sophisticated preprocessing strategies. To maintain the credibility of the dataset, various data cleaning activities were performed to remove redundancy, missing values and identify and rectify inconsistent data in structured data sets. Normalization and scaling were applied to quantify the data in order to let the machine learning algorithms work with numbers. For medical images, scaling was done as well as using flipping and rotating that can help in the expansion of the number of samples and improve the stability of a model. Matching the patient report with the right imaging data. The richness of the input is achieved with the help of feature extraction, transformation, and information augmentation while the input representation is kept interoperable across the modalities.

3.3 Model Selection and Training

In this case, an ensemble learning strategy was employed to build on the gains made by several trained AI models. Both models play a role in the general prediction of the suggested ensemble model based on XGBoost and logistic regression. XGBoost was chosen due to its parameters,like accuracy had a great potential in learning non-linear patterns, while logistic regression was selected because of its reliability and capability in learning linear patterns and probability. Recent technologies like multimodal transformers and fusion networks enable a number of kinds of data to be processed well. Depending on the type of healthcare application, the size of the data set, and CPU capability, there is always a particular model to be employed. Both the algorithms are combined using the voting technique of ensemble learning and made an effective hybrid model for prediction of diseases in the healthcare sector.

3.4 Model Evaluation

The seed set of trained ensembles has been validated using several parameters based on which the performance of the trained ensemble model has been ascertained and guaranteed. The evaluation is done on the 20% split data. Precision, accuracy, and error were used to evaluate the capability of the model in identifying positive cases and minimizing false negatives, and accuracy was used to determine the overall predictive capability of the model. Strikingly, in multimodal healthcare, significant attention is paid to the evaluation of how effectively the model fuses data from different modalities. Care delivery reliability in high-risk areas, such as health, depends on cross-validation, domain-specific benchmarks, and explainable tools..

3.5 Model Deployment

The final step of the study was the application of the ensemble model in a practical healthcare context. For the purpose of creating a scalable and accessible platform for all healthcare organizations, cloud-based architecture was chosen. During this implementation process, the model was incorporated into the extant healthcare management systems to enhance its functionality. Since experience reveals that the process of entering patient details and attaining diagnoses may not be easy sometimes, a user-friendly interface was developed. In multimodal healthcare, special care is taken to understand the measure of model integration of inputs from several modalities. External validation, domain standardization, and explainability are required techniques to attain model reliability in worldwide functioning, especially in healthcare domains.

4 Results

4.1 Accuracy

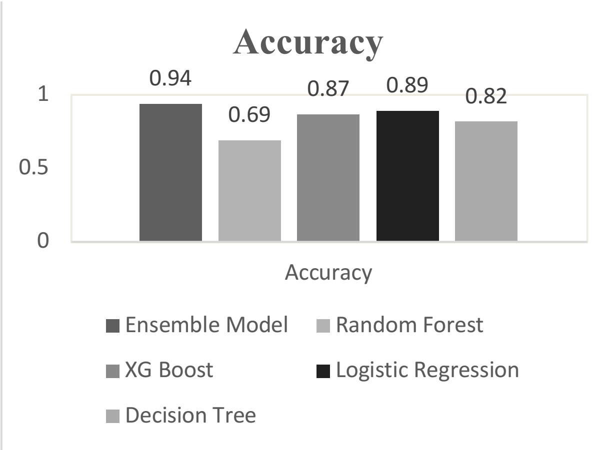

Figure 1 Accuracy of different algorithms

In the accuracy graph in Figure 1, the correctness of the total models is being compared. Accuracy means how close the expected results are to actual results. It is usually defined in terms of the number of precise predictions to the total number of forecasts made. Accuracy measures the performance of a predicted model, especially when the classes of the datasets are balanced. The performance of the ensemble model is slightly more accurate, scoring 0.94 as compared to the others. The accuracy equivalent of 0.87 and 0.89 has been achieved by XGBoost and logistic regression, respectively. The Random Forest classifier is at 0.69, and the Decision Tree classifier is slightly lower at 0.82. This reveals some of the effective features of the ensemble model, which means a number of strengths of a single algorithm can be combined for better diagnostic results.

4.2 Precision

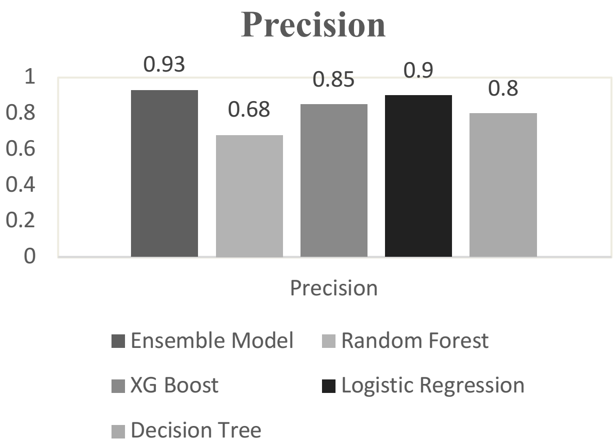

Figure 2 Precision of different ML models

The precision of the models in recognizing affirmative situations is evident from the precision in Figure 2 Precision is a performance metric that estimates the ratio of satisfactory positive predictions to the overall positive predictions of the model. It captures the degree of reliability of a model, which the model applies when it identifies positive performances. Low false positive means high accuracy because with high accuracy, false positives are rare and costly. A reduction in false positives is shown by the fact that the ensemble model has the best precision level of 0.93. Logistic regression has the precision of 0.9, followed by XGBoost 0.85, while less precision has been observed in Random Forest 0.68 and Decision Tree 0.8. This shows that the accuracy of the ensemble model becomes better than that of the single models in medical applications.

4.3 Error

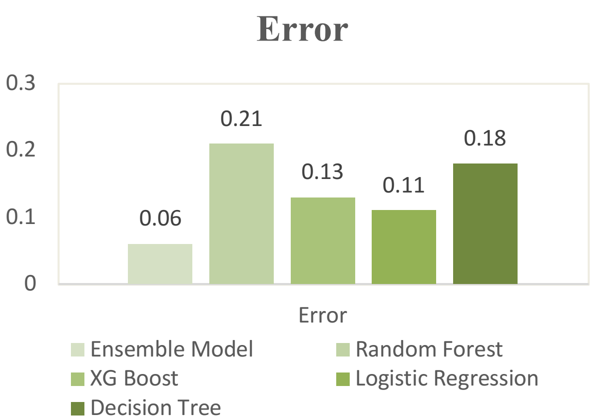

Figure 3 Errors of different ML models

From the error graph in Figure 3, where the performance of various models has been depicted. Error is what results from a difference between the expected value and the real value. It measures the discrepancy between the true value and the predictions of the model and could have been classified into bias and variance. To improve the efficiency and reliability of the predictive models, error reduction is of considerable importance. We can conclude that the ensemble model yields the lowest error rate of 0.06 and hence has better accuracy. The biggest absolute error is achieved by Random Forest, which is 0.21, then at Decision Tree: 0.18, at XGBoost: 0.13, and by logistic regression: 0.11. This shows the high performance of the ensemble model, XGBoost, and logistic regression for the classification of diseases.

5. Conclusion and Future Scope

This study, these weaknesses notwithstanding, distributed a sound healthcare diagnostic framework that was 94% accurate in the outcome with the help of a multi-modal learning approach. The model uses the important features of interpretability and prediction capability through the integration of XGBoost with logistic regression, making the model fit for most clinical applications. Due to advanced changes that help remove bias and ensure precision to make accurate predictions, the framework has the ability to deal with image data and tabular and time-series datasets. The research also showed how a comprehensive healthcare diagnostic model that involves a multimodal learning methodology achieves an exceptionally high 94% accuracy. That is why the endowed interpretability coupled with the balanced and robust predictive function of XGBoost with logistic regression makes the given model suitable for a wide range of clinical applications. Due to the high level of data pre-processing performed by the framework, it can also assess a variety of data formats like image data, tabular data, and time series data. Future advancements should focus on working out critical questions, such as how the framework can be integrated into existing health care systems and how it can be practiced in routine clinical practice. Demographic and medical variability, as well as the significance of the models, will be enhanced when there is high scalability to accommodate both minor and major datasets. Of course, to enhance its predictive performance and especially its adaptability, more investigation could consider combining ensemble learning frameworks with modal deep and conventional machine learning. One type of ethic required for constructing and adopting the model under consideration is the issue of trust and the compliance of the model with the presently existing ethical norms of AI. Also, by adding explainability to the framework, doctors will trust the use of the framework and understand projections. Emphasis on these elements will assist future advancements to ensure that the framework translates from research into application, thus increasing healthcare diagnostic accuracy and efficiency.

References

1. H. Sharma, S. Pundir, A. Deepak, K. Mayuri, S. P. Dwivedi, and N. Kumar, “Multi-Modal Data Fusion using Transfer Learning in Big Data Analytics for Healthcare,” 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI). IEEE, pp. 1–7, Dec. 29, 2023. doi: 10.1109/icaiihi57871.2023.10489309.

2. T. Zhao, L. Zhang, Y. Ma, and L. Cheng, “A Survey on Safe Multi-Modal Learning System,” 2024, arXiv. doi: 10.48550/ARXIV.2402.05355.
3. B. Chaitanya, P. N. L. Devi, S. Lakhanpal, R. B, Q. Mohammad, and B. T. Geetha, “Multi-Modal Medical Image Fusion for Enhanced Diagnosis using Deep Learning in the Cloud,” 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI). IEEE, pp. 1–7, Dec. 29, 2023. doi: 10.1109/icaiihi57871.2023.10489205.
4. M. H. Abid, R. Ashraf, T. Mahmood, and C. M. N. Faisal, “Multi-modal medical image classification using deep residual network and genetic algorithm,” PLOS ONE, vol. 18, no. 6. Public Library of Science (PLoS), p. e0287786, Jun. 29, 2023. doi: 10.1371/journal.pone.0287786.
5. A. Badhoutiya, D. P. Singh, J. R. F. Raj, A. P. Srivastava, S. L. Chari, and A. K. Khan, “Anomaly Detection in Healthcare: A Deep Learning Approach with Autoencoders,” 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI). IEEE, pp. 1–6, Dec. 29, 2023. doi: 10.1109/icaiihi57871.2023.10489790.
6. A. Abo-eleneen and A. Mohamed, “MMRL: A Multi-Modal Reinforcement Learning Technique for Energy-efficient Medical IoT Systems,” 2021 International Wireless Communications and Mobile Computing (IWCMC). IEEE, pp. 2026–2031, Jun. 28, 2021. doi: 10.1109/iwcmc51323.2021.9498842.
7. A. Ferrari, “Artificial Intelligence in Engineering and Society: Blue Skies, Black Holes, and the Job of Requirements Engineers (Keynote),” 2023 IEEE 31st International Requirements Engineering Conference Workshops (REW). IEEE, Sep. 2023. doi: 10.1109/rew57809.2023.00018.
8. Li, L., Wang, J., & Wan, S. (2024). Abstract 4800: Reducing health disparities for prostate adenocarcinoma by integrating multi-omics data via a multi-modal transfer learning approach. In Cancer Research (Vol. 84, Issue 6_Supplement, pp. 4800–4800). American Association for Cancer Research (AACR). https://doi.org/10.1158/1538-7445.am2024-4800.
9. Sturdivant, J., Morris, N., Hendricks, T., Dogan, G., &Heijnen, M. (2023). Using Artificial Intelligence to Detect Falls. In 2023 6th International Conference on Information and Computer Technologies (ICICT) (pp. 41–44). 2023 6th International Conference on Information and Computer Technologies (ICICT). IEEE. https://doi.org/10.1109/icict58900.2023.00014.
10. A. Baselizadeh, M. Z. Uddin, W. Khaksar, D. S. Lindblom, and J. Torresen, “PriMA-Care: Privacy-Preserving Multi-modal Dataset for Human Activity Recognition in Care Robots,” Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. ACM, pp. 233–237, Mar. 11, 2024. doi: 10.1145/3610978.3640701.
11. P. M. Thangaraj, S. H. Benson, E. K. Oikonomou, F. W. Asselbergs, and R. Khera, “Cardiovascular care with digital twin technology in the era of generative artificial intelligence,” European Heart Journal, vol. 45, no. 45. Oxford University Press (OUP), pp. 4808–4821, Sep. 26, 2024. doi: 10.1093/eurheartj/ehae619.
12. M. Al-Tarawneh, M. Muheilan, and Z. Al Tarawneh, “Hand Movement-Based Diabetes Detection Using Machine Learning Techniques,” International Journal on Engineering Applications (IREA), vol. 9, no. 4. Praise Worthy Prize, p. 234, Jul. 31, 2021. doi: 10.15866/irea.v9i4.20616.
13. F. A. C. Oliveira et al., “Machine Learning-Based Diabetes Detection Using Photoplethysmography Signal Features,” 2023, arXiv. doi: 10.48550/ARXIV.2308.01930.
14. P. M. Thangaraj, S. H. Benson, E. K. Oikonomou, F. W. Asselbergs, and R. Khera, “Cardiovascular care with digital twin technology in the era of generative artificial intelli-

gence," European Heart Journal, vol. 45, no. 45. Oxford University Press (OUP), pp. 4808–4821, Sep. 26, 2024. doi: 10.1093/eurheartj/ehae619.

15. R. Kumar, M. Sharma, V. Saravanan, N. Shalini, V. K. Yadav, and N. Kumar, "Graph Convolutional Networks for Disease Network Analysis in Healthcare," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI). IEEE, pp. 1–6, Dec. 29, 2023. doi: 10.1109/icaiihi57871.2023.10488947.