

**CHRIST**(DEEMED TO BE UNIVERSITY)
BANGALORE | DELHI NCR | PUNE

Notice for the PhD Viva Voce Examination

Mr Shobhit Shrotriya, Registration Number: 2071905, PhD Scholar at the Department of Statistics and Data Science, School of Sciences, CHRIST (Deemed to be University) will defend his PhD thesis at the public viva-voce examination on Saturday, 13 June 2026 at 11.30 am in Room No. 044, Ground Floor, R&D Block, CHRIST (Deemed to be University), Bengaluru - 560029, Karnataka, India.

Title of the Thesis	:	Artificial Intelligence-Based Techniques for Processing of Oncology Real-World Data
Discipline	:	Data Science
External Examiner - I	:	Dr Debi Prasanna Acharjya Professor School of Computer Sciences and Engineering Vellore Institute of Technology Tiruvalam Road, Katpadi Vellore - 632014 Tamil Nadu
External Examiner - II	:	Dr Mohammed Amjad Professor Department of Computer Engineering Jamia Millia Islamia (Central University) Jamia Nagar, Okhla New Delhi - 110025
Supervisor	:	Dr Nizar Banu P K Associate Professor Department of Computer Science School of Sciences CHRIST (Deemed to be University) Bengaluru - 560029 Karnataka

The members of the Research Advisory Committee of the Scholar, the faculty members of the Department and the School, interested experts and research scholars of all the branches of research are cordially invited to attend this open viva-voce examination.

Registrar (Academics)

Place: Bengaluru
Date: 03 June 2026

ABSTRACT

Breast cancer remains one of the most prevalent and life-threatening cancers worldwide, imposing significant challenges on healthcare systems due to its complex pathology and increasing incidence. Early detection and accurate diagnosis are essential for improving patient outcomes; however, traditional diagnostic methods and conventional artificial intelligence approach often struggle to identify subtle patterns in heterogeneous medical imaging data. Although Artificial Intelligence (AI) and Machine Learning (ML) techniques particularly Convolutional Neural Networks (CNNs)—have improved medical image analysis, they typically require large labeled datasets, are sensitive to data imbalance, and may exhibit limited generalizability across diverse imaging modalities. This research investigates the use of Large Language Model (LLM)–driven embedding techniques for breast cancer image classification and diagnostic inference. The custom-developed framework was evaluated using a real-world clinical dataset (CHBD) obtained from a leading oncology treatment center in Bangalore. The dataset consists of imaging data from 53 patients aged 32–76 years, comprising 193 multimodal medical images, including mammography (MG), positron emission tomography (PET), computed tomography (CT), and ultrasound (US) scans. To ensure robustness and external validity, the CHBD dataset was validated against five public breast imaging repositories. This cross-dataset evaluation enabled assessment of the proposed model’s ability to generalize across heterogeneous imaging modalities and dataset characteristics. Given the limited and imbalanced nature of the real-world dataset, embedding-based feature representations were employed to capture semantic similarities while reducing overfitting. Experimental results demonstrate that the LLM embedding–based approach demonstrates consistently stronger diagnostic stability than conventional CNN architectures. For the CHBD dataset, the proposed model achieved 98.44% diagnostic output consistency, compared with 61.72% obtained using CNN models, with similar performance trends observed across the public datasets. The findings suggest that embedding-driven generative AI approaches can support improved breast cancer detection, patient risk stratification, and personalized treatment planning, while emphasizing the importance of responsible AI governance for clinical adoption.

Keywords: Artificial Intelligence, Machine Learning, Generative AI, Large Language Models, Embeddings, Breast Cancer, Medical Image Analysis

Publications:

1. **Shrotriya, S.,** Banu, N. P. K., Kulkarni, A., & Aiyangar, S. (2025). Navigating the path to ethical and responsible AI integration in health and life sciences with human and machine collaboration. *Journal of the Society for Clinical Data Management*. <https://doi.org/10.47912/jscdm.412>
2. **Shrotriya, S.,** Banu, N. P. K., Kulkarni, A., & Kumar, V. G. (2025). Application of large language models for data-driven analytics in oncology: Insights and evidence generation from real-world imaging data. *International Journal of Image, Graphics and Signal Processing*, 17(6), 38–59. <https://doi.org/10.5815/ijigsp.2025.06.03>
3. **Shrotriya, S.,** Banu, P. K. N., Kulkarni, A., & Kumar, V. G. (2026). Real-world breast cancer imaging data—LLM-led analytics for insights and evidence generation. In H. Sharma, P. K. Jha, P. Jamwal, B. Tripathi, & P. Kumar (Eds.), *Proceedings of the Third Congress on Control, Robotics, and Mechatronics (CRM 2025) (Smart Innovation, Systems and Technologies, Vol. 448)*. Springer. https://doi.org/10.1007/978-981-96-8126-6_2