



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

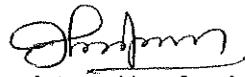
Notice for the PhD Viva Voce Examination

Ms Sonam V Maju, Registration Number: 2072901, PhD Scholar at the Department of Computer Science and Engineering, School of Engineering and Technology, CHRIST (Deemed to be University) will defend her PhD thesis at the public viva-voce examination on Saturday, 07 March 2026 at 10.15 am in CDI Conference Hall, Block V, Kengeri Campus, CHRIST (Deemed to be University), Bengaluru - 560074, Karnataka, India.

Title of the Thesis	:	Extraction and Utilization of Dark Data for the Early Prediction of Alzheimer's Disease
Discipline	:	Information Technology
External Examiner - I	:	Dr Nileshkumar Patel Associate Professor Department of Computer Science and Engineering Institute of Technology Nirma University Ahmedabad - 382481 Gujarat
External Examiner - II	:	Dr P Varalakshmi Director Centre for Artificial Intelligence and Data Science Research and Applications CEG Campus, Anna University Chennai - 600025 Tamil Nadu
Supervisor	:	Dr Gnana Prakasi O S Associate Professor Department of AI and Data Science Engineering School of Engineering and Technology Bengaluru Kengeri Campus CHRIST (Deemed to be University) Bengaluru - 560074 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.

Place: Bengaluru
Date: 24 February 2026


Registrar (Academics)

ABSTRACT

Electronic health records (EHRs) store extensive clinical information such as medical history, prescriptions, and medications. According to HIPAA regulations, these records must be stored for many years, leading to the accumulation of large volumes of data, much of which remains unused. This unused information, referred to as dark data, includes unstructured or overlooked data stored in archives, consuming storage without being analyzed. Dark data may hold untapped potential in identifying disease patterns and enabling early diagnosis. When properly extracted and analyzed, it can reveal hidden insights that improve patient outcomes and medical decision-making.

This research work develops an early prediction model for AD by discovering and using novel disease predictors that are often hidden as "dark data" in electronic health records (EHRs). The study began with an exhaustive literature review in order to determine factors previously disregarded with the onset of AD. These factors were later validated with clinical experts. The first objective of this research is to extract relevant new predictive parameters, like GlyHb, a biomarker of diabetes, which acts as dark data in the EHR. These indicators need to be identified and extracted from unstructured medical records, followed by a statistical, clinical, and technical validation to ensure the trustworthiness of the data and the clinical relevance of using these features in predictive modeling. The second objective is to identify significant AD triggering parameters by addressing dimensionality reduction issues and multicollinearity by designing a two-tier ensemble regression model for establishing feature importance and for providing strong predictive performance. Finally, in order to understand the association between the diseases, an adaptation of the Apriori algorithm and the decision tree is used to build predictive rules, and the model's validation is done mathematically, through a preclinical prediction calculator, offering a solid mathematical foundation for addressing the research gap. The methodologies used in the research work outperformed the existing methods, thereby justifying the use of dark data in enhancing the early predictions of AD.

Keywords: *Dark data, Named Entity Recognition, Ensemble regression models, Machine Learning algorithms, Mathematical model, Association rule.*

Publications:

1. **Sonam V Maju**, Gnana Prakasi O.S "A novel two-tier feature selection model for Alzheimer's disease prediction," Indonesian Journal of Electrical Engineering and Computer Science, 2024, 33(1), pp. 227–235.
2. **Sonam V Maju**, Gnana Prakasi O.S, "Mathematical Model for the Pre-Clinical Prediction of Dementia," Indian Patent, Application number: 202341005451 (2023)
3. **Sonam V Maju**, Gnana Prakasi O.S "A novel AI model for the extraction and prediction of Alzheimer's disease from electronic health records." Indonesian Journal of Electrical Engineering and Computer Science, 2025, 37(2), pp. 1023–1031.
4. **Sonam Maju & Gnana Prakasi.** ApDeC: A Rule Generator for Alzheimer's Disease Prediction. Iaes International Journal of Artificial Intelligence, 2025, 14(3), pp. 1772–1780
5. **Sonam V Maju**, Gnana Prakasi O.S., "A Dimensionality Reduction Model: A Retrospective Approach on Dementia Triggering Parameters and Feature Ranking," Communications in Computer and Information Science, 2025, 1951 CCIS, pp. 122–134
6. **Sonam V Maju**, Gnana Prakasi O.S, "Design of Decision Making Model for Integrating Dark Data from Hybrid Sectors," GRENZE International Journal of Engineering and Technology, (2022).