SAPTHAGIRI COLLEGE OF ENGINEERING

14/5, Chikkasandra, Hesaraghatta Main Road, Bengaluru – 560057.

Department of Computer Science and Engineering

Certificate

Certified that the Project Work entitled **"DISEASE PREDICTION BY MACHINE LEARNING OVER BIG DATA FROM HEALTHCARE COMMUNITIES"** carried out by **GIRISH S BATTUR (1SG15CS403), NARASIMHA M RAYKAR (1SG15CS409), SUDEEP P TULPULE (1SG15CS423), SUNIL P (1SG15CS424)**, bonafide students of **Sapthagiri College of Engineering**, in partial fulfillment for the award of **Bachelor of Engineering** in **Computer Science and Engineering** of **Visvesvaraya Technological University, Belagavi** during the academic year 2017-2018. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the department library. The project report has been approved as it satisfies the academic requirements in respect of **Project Work (10CS85)** prescribed for the said degree.

13/6/18	13/06/18	
Signature of the Guide	**Signature of the HOD**	**Signature of the Principal**
Prof. Chaithra	Dr. Yogish H K	Dr. K L Shivabasappa
Associate Professor	Professor & Head	Principal
	Professor & Head of the Department Computer Science Engineering Sapthagiri College of Engineering Bangalore-57	Principal Sapthagiri College of Engineering Chikkasandra, Hesaraghatta Road Bangalore- 560 057

EXTERNAL EXAMINATION:

Name of the Examiners	**Signature with Date**
1. MANTESH B N	13/6
2. Chaithra	13/06/18

Scanned by CamScanner

ABSTRACT

With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. A streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities experiment the modified prediction models over real-life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, a latent factor model is used to reconstruct the missing data, experiment on a regional chronic disease of cerebral infarction and propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of the knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of the proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unmoral disease risk prediction (CNN-UDRP) algorithm.

Scanned by CamScanner