An Intelligent System for Diagnosis of Pulmonary Tuberculosis Using Hybrid clustering and Random Forest algorithm

* Sıraj Sebhatu [1], Pooja[1]

[1]*Computer Science and Engineering Department, Wolaita Sodo University, Wolaita Sodo, Ethiopia*

[1*] Siraj Sebhatu, siraj.sebhatu@wsu.edu.et; Pooja, pangurala4@gmail.com

ABSTRACT

Background: Mycobacterium tuberculosis is the pathogen that causes tuberculosis, a disease that can damage the respiratory system as well as other organs of the body. It usually extends throughout the air and highest burden worldwide as one of the human infectious diseases. Especially The risk of TB attack is higher for patients with the Human Immune Deficiency Virus (HIV).In India, it is also a significant health issue Pulmonary Tuberculosis diagnosis was always a problem. Most of the detection method needs high-cost and complete power devices. In the medical field, especially in the field of PTB, due to the lack of diagnostic equipment and enough trained professionals, most patients could not get adequate services.

Objective: To design a classification and hybrid clustering model for PTB diagnosis initial screening.

Methods: The dataset was collected from 2015 to 2019, 4322 Pulmonary Tuberculosis outpatient and inpatient records were retrieved from Greater Noida Sharda University Hospital Patient History records. The Sharda University's research and ethics committee accepted this study, and informed permission was not required. Specifically, the Department of Pulmonary Tuberculosis is the source of the live data set. It is collected from the outpatient and the inpatient and is handled manually by the research team.

Results: This study introduces a model for diagnosing early pulmonary tuberculosis disease utilizing supervised and unsupervised methods. The current study successfully improved diagnostic accuracy and introduced a new hybrid model that could provide a framework for similar medical disciplines in this domain. The RandomForest model performs the high-performance testing score of AUC 97.30% using the fully optimizing trained model. Clustering-enhanced data-tuned disease diagnosis with tuned parameters and a cluster test score of 98.77.

Conclusion: The research contains a few components that require a lot of computational processing power. Particularly time-consuming is the computation of the silhouette score as well as the performance of grid search across all of the algorithms while using a wide range of parameter settings. An application of this research should emphasize parallel computing, which is not only feasible for particular segments but also algorithmic structures built on trees, like Random Forest. Clustering findings are suboptimal due to the limited number of continuous variables and the complexity of mixed clustering approaches, which led to only slight changes in the forecast. To fine-tune the algorithms, a grid search was carried out on all of the supervised algorithms to locate a set of parameters that yielded the best results in terms of the AUC score obtained from cross-validation. However, despite having somewhat lower training scores on the clustering data set, the final model that performed the best on the hold-out set was a hybrid of cluster increase and tuning.

BIOGRAPHY (100-150 words)

Siraj Sebhatu completed his PhD in Computer Science and Engineering from Sharda University. His topic of research was "Computer Recognition of Indian Sign Language". His publications include more than six papers in International Refereed Journals, presented three papers in International conferences. His research interests include Pattern Recognition, Machine Learning, Neural Networks, and Intelligent Systems. He has more than 18 years of teaching experience in various technical Institutions. He has been working at this University since November 2011. I am currently working as an Assistant Professor in the Computer Science & Engineering Department. He is also actively involved in various research and administrative activities of the University

- Mobile Number*:+251913116501
- Category*: (Oral presentation)
- Linked In
- WhatsApp No: (+251913116501)
- Research Interest*: Health Informatics
- Fax No: