Chapter 13

Sequential Attribute Designator (SAD): A Novel Feature-Selection Framework for Pulmonary Disease Research

Sulekha Das
GCECT, Kolkata, India

Avijit Kumar Chaudhuri
https://orcid.org/0000-0002-5310-3180
Brainware University, India

Debanjan Paul
Techno Engineering College, Banipur, India

Partha Ghosh
GCECT, Kolkata, India

ABSTRACT

The field of next-generation bioinformatics in pulmonary disease studies is transforming the diagnostic and treatment landscape by incorporating sophisticated algorithms for feature selection, high-dimensional data analysis, and AI-based predictive modeling. In this paper, the authors presented the Sequential Attribute Designator (SAD). This novel feature selection algorithm iteratively eliminates redundant features to reduce the feature space to the most determinative ones. SAD is an evolving form of random recursive feature elimination, which makes it different from the conventional static feature selection procedures. When used on a lung cancer dataset, SAD removed 15 features and increased classification accuracy to

DOI: 10.4018/979-8-3373-4923-7.ch013

94.11% as opposed to 78.8%. This evolution can not only minimize computational overhead but also enhance the creation of cost-effective, reliable, and interpretable diagnostic tools, thereby enabling timely interventions and improved patient survival rates in the face of pulmonary diseases.

INTRODUCTION

Lung cancer is one of the leading causes of death in the world today, and its occurrence has been on the upward trend throughout the last several decades (Thandra et al., 2021; Li et al., 2023). Though smoking accounts for 80 percent of all lung cancer occurrences, the disease can be acquired by non-smokers by way of secondary exposure to tobacco smoke or environmental pollutants like fine dust (Walser et al., 2008). The development of medical technology has allowed specialists to divide lung cancer into two main groups, namely Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC) (Pechprasarn et al., 2024; Ketkomol et al., 2024). The more common one is NSCLC, which is also broken down into squamous cell carcinoma, adenocarcinoma, and large cell carcinoma (Slatore et al., 2022). After diagnosis, other procedures include staging to ascertain the extent/level of cancer, and the most common symptoms are: dry and non-productive cough, unintentional weight loss, chronic chest pains, and fatigue (Ruano-Ravi plasma et al., 2020). Physicians can use such symptoms to diagnose the disease and prioritize the investigations to prevent needless and expensive investigations (Vidaver et al., 2016).

The diagnostic tools used to diagnose lung cancer typically include the use of imaging tools like the computed tomography (CT) scan, magnetic resonance imaging (MRI) scan, and positron emission tomography (PET) imaging. Spiral CT, along with standard chest X-ray, may be supplemented by pathological investigation, such as biopsy outcome (Cegla et al., 2023). Also, white light bronchoscopy (WLB) is usually a prevalent diagnostic method, which allows cavity pathology in trachea and bronchi to be viewed, and flexible fiberoptic bronchoscopy is the most used technique in the main line of assessing the subsegmental lobular structures of the tracheobronchial tree (Nooreldeen & Bach, 2021; Arroliga & Matthay, 1993).

Artificial Intelligence (AI) in medical diagnostics has opened up new ways of detecting lung cancer. Machine learning, expert systems, and neural networks are considered the AI technologies advanced enough to allow systems to use vast amounts of data to learn new ways of more straightforward and more accurate diagnosis (Pechprasarn et al., 2023; Sarker, 2021). SVMs are particularly suitable for use when handling structured and multidimensional data, which makes them one of the most suitable schemes to be used in diagnosing and prognosis of lung cancer (Sowmya et al., 2021). AI is also becoming a part of clinical practice, and

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/sequential-attribute-designator-sad/393269

Related Content

Protein Interactions for Functional Genomics

Pablo Minguezand Joaquin Dopazo (2012). *International Journal of Knowledge Discovery in Bioinformatics (pp. 15-30).*

www.irma-international.org/article/protein-interactions-for-functional-genomics/101240

Local Optima Avoidance in GA Biclustering using Map Reduce

Gowri R.and Rathipriya R. (2016). *International Journal of Knowledge Discovery in Bioinformatics (pp. 37-47).*

www.irma-international.org/article/local-optima-avoidance-in-ga-biclustering-using-map-reduce/171418

Molecular Network Analysis of Target RNAs and Interacting Proteins of TDP-43, a Causative Gene for the Neurodegenerative Diseases ALS/FTLD

Jun-Ichi Satoh (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications (pp. 964-985).*

www.irma-international.org/chapter/molecular-network-analysis-target-rnas/76105

Medical Domain Knowledge and Associative Classification Rules in Diagnosis

Sung Ho Ha (2011). *International Journal of Knowledge Discovery in Bioinformatics (pp. 60-73).*

www.irma-international.org/article/medical-domain-knowledge-associative-classification/52771

The Avatar as a Self-Representation Model for Expressive and Intelligent Driven Visualizations in Immersive Virtual Worlds: A Background to Understand Online Identity Formation, Selfhood, and Virtual Interactions

Colina Demirdjianand Hripsime Demirdjian (2020). *International Journal of Applied Research in Bioinformatics (pp. 1-9).*

www.irma-international.org/article/the-avatar-as-a-self-representation-model-for-expressive-and-intelligent-driven-visualizations-in-immersive-virtual-worlds/261865