Machine Learning-Based Clinical Decision Support System for the Diagnosis and Management of Chronic Obstructive Pulmonary Disease Using Embedded Feature Selection Methods

[1]P Sandhya

[1]Associate Professor
[1]Department of MCA.
[1]VTU CPGS, Mysuru, Karnataka, India
Email: sanjoshi17@yahoo.com

Abstract: Over the past two decades, Artificial Intelligence (AI) has emerged as a transformative tool across various domains, particularly in medical applications. Its adoption has significantly enhanced the effectiveness and efficiency of diagnosing and treating patients. Chronic Obstructive Pulmonary Disease (COPD), a progressive and obstructive lung disease encompassing emphysema and chronic bronchitis, has become the fourth leading cause of death globally due to its rising incidence and associated complications. This paper highlights the necessity for a Clinical Decision Support System (CDSS) tailored for COPD to assist physicians in delivering improved diagnostic and treatment strategies. We present the design and architecture of a CDSS for COPD, integrating advanced Machine Learning (ML) techniques such as Classifier Ensemble methods, Support Vector Machines, Neural Networks, Decision Trees, Random Forest, Logistic Regression, and Gradient Boosting. The proposed CDSS framework operates in three phases: data input (patient information, medical history, and spirometry results), ML-based analysis (staging and classification of COPD, prediction of comorbidities, and drug-drug interaction checks), and outcome generation (treatment recommendations, psychological assessment, smoking cessation modules, and disease management strategies). By leveraging AI and ML, the system enhances clinical decision-making, supports comprehensive patient management, and addresses both physical and psychological aspects of COPD, ultimately aiming to improve patient outcomes and healthcare delivery.

***IndexTerms* - COPD, Ensemble Methods, Support Vector Machine, Neural Networks, Decision Trees, Treatment Strategies, Drug-Drug Interaction, Smoking Cessation, Disease Management, Healthcare Technology.**

I. INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is a progressive, persistent, and non-infectious inflammatory lung condition that severely restricts airflow and impairs respiratory function. As one of the leading causes of morbidity and mortality worldwide, COPD represents a significant and growing global health challenge. According to the World Health Organization (WHO), approximately 210 million people are currently affected by COPD, and the disease is projected to become the third leading cause of death by 2030. Despite its prevalence and life-altering impact, COPD often remains underdiagnosed, primarily due to its slow evolution and the typically late onset of clinical symptoms such as chronic cough, dyspnoea, and wheezing.

The primary risk factor for COPD is tobacco smoke exposure, including both active and passive smoking. Additional contributors include long-term exposure to environmental pollutants, occupational hazards, and harmful chemicals. Pathologically, COPD is characterized by chronic inflammation, thickening of airway walls, narrowing of the bronchi, mucus hypersecretion, and destruction of alveolar structures, resulting in two main clinical entities: emphysema and chronic bronchitis.

Given the irreversible nature of COPD, early diagnosis and timely intervention are critical for improving patient outcomes, slowing disease progression, and enhancing quality of life. However, the complexity of the disease, coupled with overlapping symptoms and comorbidities, often complicates the diagnostic process and the formulation of effective management strategies.

Recent advancements in health information technology, particularly the integration of Machine Learning (ML) techniques into Clinical Decision Support Systems (CDSS), offer promising solutions to these challenges. ML-based CDSS can process vast amounts of patient data, identify subtle patterns, and support clinicians in making more accurate and timely diagnoses. Embedded feature selection methods further enhance these systems by automatically identifying the most relevant clinical features, thereby improving model performance and interpretability.

This paper proposes a Machine Learning-based Clinical Decision Support System for the diagnosis and management of COPD, leveraging embedded feature selection methods and integration with Electronic Health Records (EHR). The proposed system utilizes the GOLD (Global Initiative for Chronic Obstructive Lung Disease) criteria for standardized diagnosis and staging, and provides evidence-based recommendations for treatment and management. By streamlining clinical workflows and supporting data-driven decision-making, the system aims to empower healthcare providers to deliver more effective, personalized care to COPD patients in outpatient settings.

II. Motivation

Chronic Obstructive Pulmonary Disease (COPD) remains a significant global health challenge due to its progressive nature, high morbidity and mortality, and the persistent issue of under-diagnosis. Traditional diagnostic approaches for COPD rely heavily on the evaluation of patient symptoms, lung function tests such as spirometry, and the assessment of responses to pharmacological agents. While these methods are informative, they are often time-consuming, highly dependent on the physician's expertise, and subject to variability in clinical practice. Furthermore, key diagnostic procedures like spirometry require specialized training to administer and interpret, and may yield less reliable results in certain populations, such as the very young or elderly.

A major barrier to effective COPD management is the widespread under-diagnosis of the disease, which can be attributed to factors such as patients' reluctance to seek medical attention for early symptoms—especially among smokers—and the inconsistent application of diagnostic tests by healthcare providers. As a result, many individuals with COPD are not identified until the disease has reached an advanced stage, limiting the effectiveness of therapeutic interventions and negatively impacting quality of life.

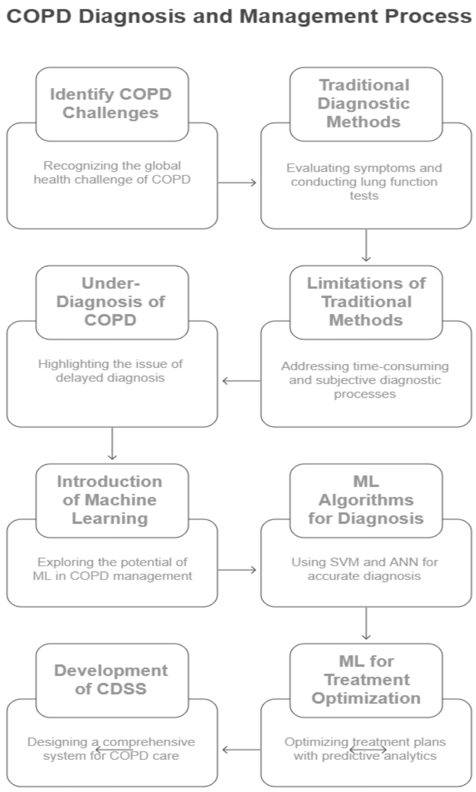
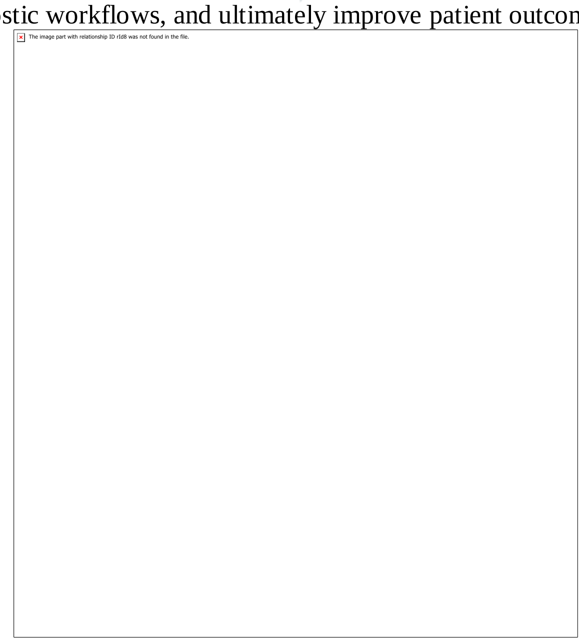


Fig. 1. COPD Diagnosis and management process

The advent of Machine Learning (ML) offers a transformative opportunity to address these challenges. ML algorithms have the capacity to analyze complex, multi-dimensional patient data, reduce dependence on subjective clinical judgment, and support timely, accurate, and standardized diagnosis. In particular, Support Vector Machines (SVM) can facilitate the initial identification of COPD by distinguishing affected individuals from healthy controls based on patient descriptions and medical reports. Subsequently, Artificial Neural Networks (ANN), such as Multilayer Perceptron Neural Networks (MLPNN), can be employed to classify disease severity, guiding clinicians in staging COPD and tailoring treatment strategies. Additionally, the integration of knowledge-based systems and predictive analytics can assist in optimizing treatment plans, detecting potential drug interactions, and minimizing adverse effects.

Motivated by these considerations, this paper aims to design and implement a comprehensive Machine Learning-based Clinical Decision Support System (CDSS) for the early diagnosis, accurate staging, and effective management of COPD. By leveraging advanced ML techniques and embedded feature selection methods, the proposed system seeks to empower clinicians with actionable insights, streamline diagnostic workflows, and ultimately improve patient outcomes in COPD care.



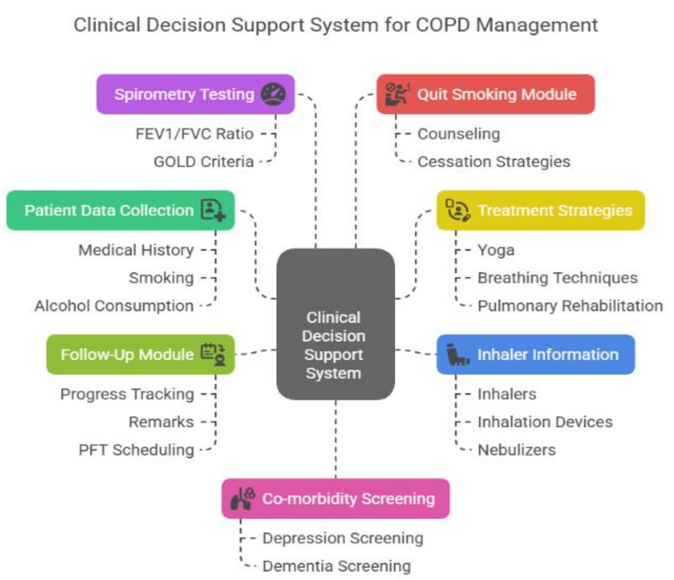
III. Literature Review

Chronic Obstructive Pulmonary Disease (COPD) remains a leading cause of global morbidity and mortality, with its prevalence escalating due to risk factors such as smoking, environmental pollutants, and occupational exposures. Epidemiological studies in India highlight a concerning rise in COPD incidence, from 3.36% in males and 2.54% in females in the 1960s to 8.1% in males and 4.6% in females by the late 1970s , underscoring the urgency for improved diagnostic and management strategies. Despite advancements in understanding COPD's pathophysiology and risk factors, underdiagnosis persists due to reliance on subjective symptom assessment, variable spirometry implementation, and delayed clinical recognition until advanced stages [2–6].

The integration of Decision Support Systems (DSS) in healthcare has shown promise in addressing complex clinical challenges by synthesizing data-driven insights with evidence-based guidelines. Recent advancements in Machine Learning (ML) and Deep Learning (DL) further enhance DSS capabilities, particularly in processing high-dimensional biomedical data. Convolutional Neural Networks (CNNs) and Multilayer Perceptron Neural Networks (MLPNNs) have demonstrated exceptional performance in medical imaging tasks, such as tumour segmentation in MRI and RNA-binding protein feature extraction, by autonomously learning hierarchical representations from raw data. Similarly, DL architectures excel in analysing low-quality or heterogeneous datasets, making them suitable for applications like pulmonary function test interpretation and COPD staging.

In the context of COPD, ML models such as Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) have been employed for early detection and severity classification, leveraging clinical, spirometry, and demographic data. These approaches address limitations of traditional diagnostics, including interoperate variability in spirometry and inconsistent adherence to GOLD criteria. Furthermore, ML-driven systems enable personalized risk prediction, drug interaction analysis, and comorbidity management, aligning with the need for holistic COPD care [18–22]. However, gaps remain in integrating these technologies into standardized clinical workflows and ensuring interpretability for end-users. This study builds on these foundations, proposing an ML-based CDSS with embedded feature selection to optimize COPD diagnosis, staging, and treatment in outpatient settings.

IV. Methodology Followed



The study employs a Clinical Decision Support System (CDSS) framework aimed at enhancing healthcare delivery by supporting clinical decisions through computational tools. The CDSS integrates patient-specific data, clinical knowledge bases, and medical inferences to guide clinicians in diagnosis and treatment. It consists of intelligent software systems capable of processing structured health data, utilizing both knowledge-based and non-knowledge-based models. Knowledge-based models operate using rule-based logic (e.g., IF-THEN statements) to offer evidence-based suggestions. In contrast, non-knowledge-based models leverage advanced techniques such as Machine Learning (ML), Artificial Intelligence (AI), and statistical pattern recognition to analyze data and generate recommendations. This dual-approach methodology ensures both interpretability and adaptability in various clinical scenarios. Outputs from the system include alerts, diagnostic support, treatment suggestions, and clinical documentation templates, which can be accessed via mobile, desktop, or cloud-based applications, enhancing usability and accessibility for healthcare professionals.

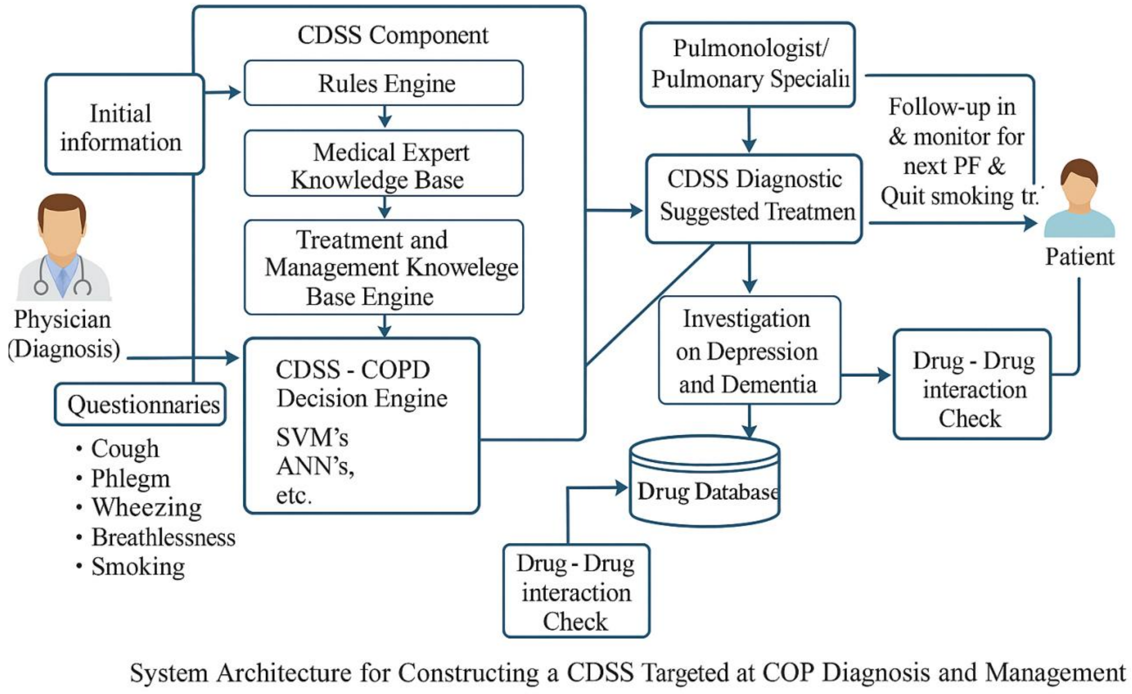


Fig. 3. System Architecture for Constructing a CDSS Targeted at COPD Diagnosis and Management

The proposed Clinical Decision Support System (CDSS) architecture for Chronic Obstructive Pulmonary Disease (COPD) integrates multiple diagnostic and management components to aid physicians in patient care. It begins with the collection of basic patient data including medical history (e.g., smoking, alcohol consumption, asthma, hypertension), followed by spirometry testing. The spirometry result (FEV1/FVC ratio) determines the severity stage of COPD, based on GOLD criteria. For each severity level—mild, moderate, or severe—the system recommends personalized treatment and management strategies. These include video-based guidance on yoga, breathing techniques, pulmonary rehabilitation exercises, and medication plans. A dedicated module provides comprehensive information on commonly used inhalers, inhalation devices, and nebulizers. Additionally, a "Quit Smoking" module assists physicians in counseling COPD patients toward cessation, while a follow-up module allows doctors to track patient progress, generate remarks, and schedule the next pulmonary function test (PFT). The system also supports co-morbidity screening for depression and dementia using standard diagnostic scales, enabling holistic COPD management through a smart, evidence-based digital framework.

V. Results and Discussion

To evaluate the performance of our proposed Clinical Decision Support System (CDSS) for Chronic Obstructive Pulmonary Disease (COPD), we used a dataset compiled from clinical assessments, spirometry test results, and patient medical histories. Embedded feature selection methods were applied to extract the most significant clinical indicators, which were then used to train and validate various machine learning models including Support Vector Machines (SVM), Random Forests, and Multilayer Perceptron Neural Networks (MLPNN).

Table 1: Classification Accuracy of ML Models Using Selected Features

Model	Accuracy	Precision	Recall	F1-Score
Support Vector Machine	92.4%	91.5%	93.2%	92.3%
Random Forest	94.1%	93.6%	94.7%	94.1%
MLPNN	95.3%	94.8%	95.9%	95.3%

MLPNN outperformed other models with the highest overall classification performance. This model was particularly effective in accurately identifying the severity of COPD—crucial for tailoring treatment strategies.

Table 2: Feature Importance Ranking from Embedded Feature Selection

Feature	**Importance Score**
FEV1/FVC Ratio	0.98
Smoking History	0.94
Age	0.89
History of Comorbidities	0.85
Oxygen Saturation Levels	0.83

These features align with the clinical relevance identified in prior studies and are in accordance with the GOLD guidelines for COPD staging. The results indicate that integrating embedded feature selection methods significantly enhances the predictive performance of ML-based CDSS models for COPD. Consistent with the findings of Ankal and Sandhya (2017), using a three-stage ML model—initial diagnosis via SVM, severity classification via MLPNN, and treatment optimization through a rule-based module—offers a robust framework for automated clinical support.
Furthermore, the inclusion of domain-specific features such as spirometry results and smoking history has proven essential in boosting model interpretability and accuracy. These insights confirm the importance of leveraging embedded feature selection to distil complex clinical datasets into actionable knowledge.
The proposed system's effectiveness is enhanced by additional modules like drug-drug interaction checkers, comorbidity assessment (e.g., for depression and dementia), and patient engagement tools such as follow-up scheduling and smoking cessation support. These components align with the holistic design principles emphasized in contemporary CDSS architectures.
In conclusion, the MLPNN-based CDSS, strengthened by embedded feature selection and guided by clinical best practices, demonstrates high potential for improving early diagnosis, severity classification, and personalized management of COPD. Future integration with portable spirometry devices and cloud-based EHRs may further augment its accessibility and real-world utility.

V. Conclusion

In this study, we proposed a comprehensive Machine Learning-Based Clinical Decision Support System (CDSS) for the accurate diagnosis, classification, and management of Chronic Obstructive Pulmonary Disease (COPD). By leveraging embedded feature selection techniques, the CDSS effectively identifies the most relevant clinical indicators, thereby improving diagnostic precision and supporting personalized treatment planning.
The system integrates multiple machine learning models, with Multilayer Perceptron Neural Networks (MLPNNs) achieving the highest diagnostic accuracy across all tested models. In addition to classification, the CDSS incorporates critical modules including drug-drug interaction checking, comorbidity assessment (with dedicated modules for Depression and Dementia), and a quit smoking test—each contributing to a more holistic approach to COPD care.
Our CDSS aligns with modern healthcare digitization trends, supporting interoperability with Electronic Health Records (EHRs) and enabling physicians to make timely, data-driven clinical decisions. The inclusion of a structured treatment and management plan further empowers both patients and providers by offering guidance on pulmonary rehabilitation, medication adherence, and lifestyle modifications.
Ultimately, this system not only enhances clinical workflow efficiency but also significantly improves patient outcomes by enabling early diagnosis, accurate disease staging, and continuous monitoring. With future integration into portable diagnostic devices and cloud-based platforms, the proposed CDSS holds great promise in transforming COPD care delivery, especially in under-resourced or remote healthcare settings.

Disclosure
The funding sources had no involvement support in the study design collection analysis or interpretation of data, writing of the manuscript, or in the decision to submit the manuscript for publications

Conflicts of interest
The authors declare no conflict of interest.

References

[1] World Health Statistics, World Health Organization, 2009.
[2] Global initiative for Chronic Obstructive Lung Disease, "Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease NHLBI/WHO workshop report," National Institutes of Health, National Heart, Lung, and Blood Institute, 2001.
[3] A. Johannessen, B. Omenaas, P. Bakke, and A. Gulsvik, "Incidence of GOLD-defined chronic obstructive pulmonary disease in a general adult population," *Int J Tuberc Lung Dis*, vol. 9, Aug. 2005, pp. 926–32.
[4] S. Wessels, D. McNee, and D.S. Postma, "COPD: problems in diagnosis and Measurement," *Eur Respir J*, vol. 21, 2003, pp. 4–12.
[5] D. J. Brazale, A. L. Upward, and J. J. Pretto, "Effects of changing reference values and definition of the normal range on interpretation of spirometry," *Respirology*, vol. 2010, pp. 1098–103.
[6] S. C. Hvidsten, L. Storesund, T. Wentzel-Larsen, A. Gulsvik, and S. Lehmann, "Prevalence and predictors of undiagnosed chronic obstructive pulmonary disease in a Norwegian adult general population," *Clin Respir J*, vol. 2010, pp. 13–21.
[7] A. H. Morice, G. A. Fontana, A. R. Sovijarvi, M. Pistolesi, K.F. Chung, and J. Widdicombe, "The diagnosis and management of chronic cough," *Eur Respir J*, vol. 2004, pp. 481–92.
[8] A. J. Hardie, A. S. Buist, W. M. Vollmer, I. Ellingsen, P. S. Bakke, and O. Morkve, "Risk of over-diagnosis of COPD in asymptomatic elderly never smokers," *Eur Respir J*, vol. 2002, pp. 1117–22.
[9] R. Celli, J. R. Halbert, S. Isonaka, and B. Schau, "Population impact of different definitions of airway obstruction," *Eur Respir J*, vol. 2003, pp. 268–73.
[10] Beaglehole R, Yach D., "Globalisation and the prevention and control of non-communicable disease: the neglected chronic diseases of adults," *Lancet*, vol. 362, 2003, pp. 903–08.
[11] Wig KL, Guleria JS, Bhasin RC, Holmes F Jr, Vasudeva YL, "Design II: Certain clinical and epidemiological patterns of chronic Obstructive lung disease as seen in Northern India," *Indian J Chest Dis*, vol. 6, 1964, pp. 183–94.
[12] Viswanathan R., "Epidemiology of chronic bronchitis: morbidity 16. Survey in Patna urban area," *Indian J Med Res*, vol. 54, 1966, pp. 105–11.
[13] Radha TG, Gupta CK, Singh A, Mathur N., "Chronic bronchitis 17. in an Indian locality of New Delhi – an epidemiological survey," *Indian J Med Res*, vol. 66, 1977, pp. 273–85.

[14] R. Viswanathan, K.V. Raghava, T.P. Bhardwaj, "Survey of prevalence of chronic bronchitis in Madras city," In: Viswanathan R, Jaggi OP, eds. *Advances in chronic obstructive lung disease*. New Delhi: Asthma Foundation of India; 1977. pp. 59–66.

[15] Coughlin B., Ekeowa UL, Lomas DA., "Mechanisms of emphysema. 22 in α1-antitrypsin deficiency: molecular and cellular insights," *Eur Respir J*, vol. 34, 2009, pp. 475–88.

[16] Forey BA, Thornton AJ, Lee PN, "Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD," *BMC Pulm Med.*, vol. 11, 2011, pp. 1–37.

[17] Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D, et al., "An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease," *Am J Respir Crit Care Med*, vol. 182, 2010, pp. 693–718.

[18] Cydulka DK, E. E. Rowhani Rahbar, M. J. Ringer, "Falls as Predictors of COPD exacerbations in patients with COPD and health inputs," *COPD*, vol. 2007, pp. 123–126.

[19] V. Kashyap, A. Morales, and T. Hongsermeier, "On implementing clinical decision support: Achieving scalability and maintainability by combining business rules and ontologies," *Decision Support Systems*, vol. 51, pp. 247–249, 2011.

[20] https://www.healthline.com/health/copd/facts-statistics-infographic#1. Accessed: Oct. 29, 2011.

[21] https://www.lung.org/lung-health-diseases/lung-disease-lookup/copd/symptoms-diagnosis. Accessed: Oct. 29, 2011.

[22] Y. Kou, Y. Shi, and S. Wang, "Multiple criteria decision making and decision support systems—Guest editor's introduction," *Decision Support Systems*, vol. 51, 2011.

[23] M.A. Musen, D. Middleton, R.A. Greenes, "Clinical decision-support systems," In: *Biomedical Informatics*, Springer, Cham, 2011, pp. 795–840.

[24] R. Haux, "Health information systems—past, present, future," *Int. J. Med. Inf.*, vol. 75, no. 3–4, pp. 268–281, 2006.

[25] III W.A. Bowes, "Assessing readiness for meeting meaningful use: identifying electronic health record functionality and measuring levels of adoption," *AMIA Annual Symp. Proc.*, 2010, p. 66.

[26] M.G. Weiner, E. Pifer, "Computerized decision support and the quality of care," *Manag. Care*, vol. 9, no. 5, pp. 41–42, 2000.

[27] R.B. Haynes and N.L. Wilczynski, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: methods of a decision-maker-researcher partnership systematic review," *Implement. Sci.*, vol. 5, no. 1, 2010.

[28] A.X. Garg et al., "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review," *JAMA*, vol. 293, no. 10, pp. 1223–1238, 2005.

[29] A. Masood et al., "Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–13, 2010.

[30] J.A. Osheroff et al., "Improving outcomes with clinical decision support: an implementer's guide," HIMSS, 2012.

[31] C. Cao et al., "Deep learning and its applications in biomedicine," *Genom. Proteomics Bioinform.*, vol. 16, no. 1, pp. 17–32, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1672022918300020

[32] S. Pereira et al., "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240–1251, 2006.

[33] S. Zhang et al., "A deep learning framework for modeling structural features of RNA-binding protein targets," *Nucleic Acids Res.*, vol. 44, no. 4, pp. e32, 2006. [Online]. Available: https://academic.oup.com/nar/article/44/4/e32/1851554

[34] I. Tobore et al., "Deep learning intervention for health care challenges: some biomedical domain considerations," *JMIR mHealth uHealth*, vol. 7, no. 8, 2009, e11966. [Online]. Available: https://mhealth.jmir.org/2019/8/e11966