iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: II Month of publication: February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77547

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Patient Case Similarity using AI with Black Fungus and Cancer Treatment

Anand Pal[1], Alok Singh[2], Abhinav Pandey[3], Ms. Shekhar Srivastav[4]
[1, 2, 3]Dept. Computer Science & Engineering Babu Banarasi Das Institue of Technology & Management (Dr A P J Abdul Kalam Technical University) Lucknow, India
[4]Assistant Professor, Dept. Computer Science & Engineering Babu Banarasi Das Institue of Technology & Management (Dr A P J Abdul Kalam Technical University) Lucknow, India

Abstract: This paper presents an Artificial Intelligence (AI)-based Patient Case Similarity System designed to assist doctors in analyzing and comparing cases of black fungus (mucormycosis) and cancer patients. The system uses machine learning algorithms to identify similar historical patient cases based on symptoms, laboratory reports, medical imaging features, comorbidities, and treatment responses. By applying similarity measures and classification techniques such as Random Forest and Cosine Similarity, the system helps clinicians make faster and evidence-based treatment decisions. The proposed model improves diagnostic support, enhances personalized treatment planning, and reduces clinical uncertainty, especially in high-risk conditions like mucormycosis and oncology cases.
Keywords: Artificial Intelligence, Patient Similarity, Black Fungus, Cancer Treatment, Machine Learning, Clinical Decision Support

I. INTRODUCTION

Healthcare systems worldwide are experiencing a rapid digital transformation due to the increasing availability of electronic health records (EHRs), medical imaging data, laboratory reports, and genomic information. This vast amount of clinical data provides an opportunity to apply Artificial Intelligence (AI) techniques to improve diagnosis, prognosis, and personalized treatment planning. Traditional clinical decision-making relies heavily on physician expertise and manual analysis of patient history, which may become challenging when dealing with complex and high-risk diseases such as cancer and mucormycosis.

One such life-threatening condition is Mucormycosis, commonly known as black fungus. Mucormycosis is a rare but aggressive fungal infection caused by fungi belonging to the Mucorales order. It primarily affects individuals with weakened immune systems, including diabetic patients, cancer patients undergoing chemotherapy, organ transplant recipients, and patients treated with high-dose steroids. During the COVID-19 pandemic, India reported a significant surge in mucormycosis cases, particularly among patients with uncontrolled diabetes and prolonged steroid therapy. The infection can rapidly spread from the sinuses to the eyes and brain, leading to severe complications and high mortality rates if not diagnosed and treated promptly. Early identification of similar historical cases can significantly improve clinical outcomes.

Parallelly, Cancer remains one of the leading causes of death globally. Cancer is characterized by uncontrolled cell growth and the potential to invade or spread to /other parts of the body (metastasis). Treatment strategies for cancer are highly personalized and depend on multiple factors such as tumor stage, histopathology findings, genetic mutations, patient age, comorbidities, and previous treatment response. Oncologists must analyze large volumes of patient data to determine the most effective therapeutic approach, including chemotherapy, radiotherapy, immunotherapy, targeted therapy, or surgical intervention. However, manually comparing a new patient’s case with thousands of previous records is both time-consuming and prone to human limitations.Artificial Intelligence, particularly Machine Learning (ML), offers a powerful solution to this challenge. AI algorithms can process multidimensional clinical data and identify hidden patterns that may not be visible through traditional statistical methods. One promising approach in precision medicine is the concept of Patient Case Similarity, where a new patient’s clinical profile is compared with historical patient data to identify the most similar cases. This approach enables clinicians to learn from previous treatment outcomes and apply evidence-based strategies tailored to the individual patient.

Patient Similarity Models typically represent each patient as a feature vector containing demographic details, clinical symptoms, laboratory values, imaging findings, disease stage, and treatment history. Similarity metrics such as Cosine Similarity, Euclidean Distance, or K-Nearest Neighbors (KNN) are then used to measure the closeness between patient records. Machine learning algorithms like Random Forest, Support Vector Machines (SVM), and Deep Learning models further enhance prediction accuracy and decision support capabilities. In the context of mucormycosis, patient similarity analysis can help identify high-risk individuals based on blood glucose levels, steroid exposure, oxygen therapy history, and CT scan findings. For cancer patients, similarity analysis can assist in predicting survival rates, treatment response, recurrence probability, and optimal therapy combinations. By leveraging historical clinical data, AI-based systems can reduce diagnostic delays, improve treatment selection, and enhance overall patient survival.

The objective of this research is to develop an AI-driven Patient Case Similarity System for black fungus and cancer treatment. The proposed system integrates data preprocessing, machine learning-based prediction, and similarity computation to retrieve the most relevant historical cases. This approach

II. LITERATURE SURVEY

The concept of patient case similarity has gained significant attention in recent years due to the rapid growth of Electronic Health Records (EHRs) and the increasing demand for personalized medicine. Artificial Intelligence (AI) and Machine Learning (ML) techniques have been widely adopted to analyze patient data, identify patterns, and support clinical decision-making.

A. Patient Similarity Modeling in Healthcare

Patient similarity modeling aims to measure how closely two patients resemble each other based on clinical, demographic, laboratory, and imaging data. Early approaches relied on distance-based statistical measures such as Euclidean distance and Manhattan distance to compare structured patient attributes. However, these methods struggled with high-dimensional and heterogeneous healthcare data.Recent studies have introduced advanced representation learning techniques to generate patient embeddings from EHR data. Deep learning models such as autoencoders and recurrent neural networks have been used to capture temporal patterns in longitudinal health records. These embedding-based methods improve similarity search performance and allow more meaningful clustering of patients with similar disease progressionOntology-based similarity frameworks have also been proposed to enhance semantic understanding of medical concepts. By mapping clinical terms to standardized vocabularies, similarity calculations become more interpretable and clinically meaningful.

B. Machine Learning in Cancer Diagnosis and Treatment

Artificial intelligence has shown significant success in cancer detection, prognosis, and treatment recommendation. Supervised learning algorithms such as Support Vector Machines (SVM), Random Forest, and Gradient Boosting have been widely used for cancer classification tasks. Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated high accuracy in analyzing medical imaging data including CT scans, MRI scans, and histopathology slides. These models automatically extract tumor-related features such as shape, size, and texture. Furthermore, survival analysis techniques combined with machine learning models have been applied to predict patient survival rates and chemotherapy responses. Patient similarity networks have been developed to group cancer patients into clinically meaningful cohorts, enabling personalized treatment planning.

Despite these advancements, many existing cancer prediction systems focus primarily on classification accuracy rather than similarity-based case retrieval, which is essential for treatment comparison.

C. AI Applications in Mucormycosis (Black Fungus)

Mucormycosis, commonly known as Black Fungus, emerged as a serious opportunistic infection during the COVID-19 pandemic, especially among diabetic and immunocompromised individuals. Mucormycosis is characterized by rapid tissue invasion, sinus involvement, and high mortality rates if not treated promptly. Most research on Mucormycosis focuses on clinical characteristics, risk factors, and antifungal treatment protocols. Studies identify uncontrolled diabetes, steroid usage, and oxygen therapy as major contributing factors. However, limited research exists on AI-driven case similarity modeling for fungal infections.

Compared to cancer research, AI applications in fungal infection analytics remain underdeveloped. This creates an opportunity for building intelligent similarity-based systems to assist clinicians in early detection and treatment strategy selection.

D. Similarity Metrics and Case-Based Reasoning

Similarity computation plays a critical role in patient case retrieval systems. Various similarity measures have been studied in healthcare analytics:

Euclidean Distance – Suitable for continuous numerical attributes

Cosine Similarity – Effective for high-dimensional feature vectors

Jaccard Similarity – Used for binary symptom matching

Mahalanobis Distance – Accounts for feature correlation

Case-Based Reasoning (CBR) systems use similarity scores to retrieve past cases and adapt their solutions to new problems. In healthcare, CBR has been applied to treatment planning and diagnosis support. However, traditional CBR systems lack integration with modern deep learning feature extraction techniques.

Recent research suggests that combining machine learning embeddings with similarity metrics significantly improves retrieval performance and clinical relevance.

E. Research Gaps

Although substantial progress has been made in AI-driven healthcare analytics, several limitations remain:

Most systems focus on single disease classification rather than multi-disease similarity comparison.

Limited integration of structured clinical data and imaging features in unified similarity frameworks.

Scarcity of AI-based similarity research specifically targeting Mucormycosis.

Lack of interpretable models that support clinician trust and explainability.

These gaps motivate the development of a unified Patient Case Similarity System that integrates clinical, laboratory, and imaging features to support treatment decision-making for both Mucormycosis and Cancer.

III. PROPOSED SYSTEM AND METHODOLOGY

The rapid growth of Electronic Health Records (EHRs) has enabled the development of AI-driven clinical decision support systems [1], [2]. Deep learning and machine learning techniques have demonstrated strong potential in extracting meaningful patterns from structured and unstructured healthcare data [3], [20].

Patient similarity modeling is an important research area in precision medicine. Recent studies have introduced patient similarity networks to group clinically comparable cases using EHR embeddings [4], [19]. Embedding-based models improve retrieval of similar cases compared to traditional statistical approaches [2]. Distance metrics such as Euclidean distance, cosine similarity, and Mahalanobis distance have been evaluated for clinical similarity search [6].

In cancer research, artificial intelligence has been widely applied for tumor detection, classification, and prognosis prediction [7], [8], [9]. Convolutional Neural Networks (CNNs) have achieved dermatologist-level performance in image-based cancer classification [7]. Ensemble learning models such as Random Forest have also shown strong predictive performance in clinical datasets [15].

Transfer learning approaches further improve prediction accuracy when training data is limited [10]. Deep learning architectures applied to EHR systems enhance risk prediction and treatment outcome estimation [3], [20].

Regarding fungal infections, particularly Mucormycosis, clinical studies have primarily focused on epidemiology, diagnosis, and antifungal therapy [11], [12], [13]. However, limited research exists on AI-based similarity modeling for mucormycosis cases. Most current studies concentrate on clinical characteristics rather than computational case retrieval systems.

Case-Based Reasoning (CBR) systems use similarity computation to retrieve past cases and adapt solutions to new problems [21]. Nearest neighbor-based pattern classification remains one of the foundational approaches in similarity modeling [16]. Modern AI-based clinical decision systems extend these approaches with high-dimensional feature learning [24].

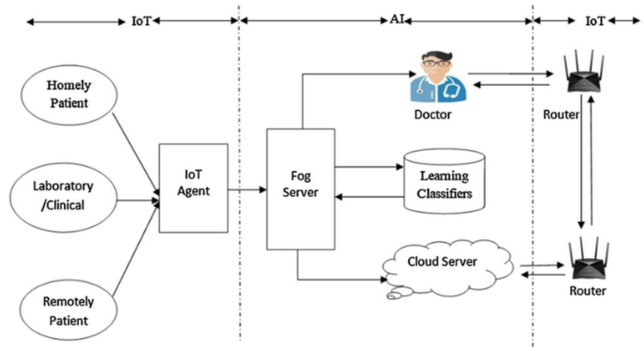
Although significant progress has been made in cancer analytics [9], [20], there remains a gap in unified similarity-based systems integrating both cancer and mucormycosis cases. The proposed research addresses this gap by developing a multi-disease similarity framework.

A. System Architecture Overview

The proposed system consists of five major modules:

1) Data Acquisition
2) Preprocessing and Feature Engineering
3) Patient Vector Representation
4) Similarity Computation Engine
5) Treatment Recommendation Module

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com



B. Data Acquisition and Preprocessing

Clinical data is collected from EHR systems and public repositories [17]. Features for mucormycosis include blood glucose level, steroid exposure, and infection severity [11]. Cancer-related features include tumor size, TNM stage, and histopathology results [9].
Preprocessing steps include:

- Missing value imputation
- Normalization
- Categorical encoding
- Imaging feature extraction using CNN-based encoders [8]

C. Preprocessing Pipeline

Since raw medical data is complex, preprocessing is critical.

Features Used:

Category	Example Features
Demographics	Age, Gender
Clinical	Fever, Pain, Swelling
Lab Reports	Blood Sugar, WBC Count
Imaging	CT/MRI lesion size
Treatment	Chemotherapy, Amphotericin-B
Outcome	Recovered / Complication / Death

Preprocessing Steps:
Step 1: Remove missing or corrupted patient records
Step 2: Encode categorical variables (Gender, Cancer type)
Step 3: Normalize lab values
Step 4: Convert medical text reports using NLP (TF-IDF / BERT embeddings)
Step 5: Extract image features using CNN models
Step 6: Combine structured + unstructured features
Keeping only clinically relevant features improves similarity accuracy and avoids noise in prediction.

D. Machine Learning Model for Patient Case Similarity

Instead of regression (like FertiCast), your system will use:
1. Similarity-Based Models
K-Nearest Neighbors (KNN)
Cosine Similarity
Eucidean Distance

Siamese Neural Networks
Transformer-based medical embeddings
Proposed Algorithm: Hybrid Similarity Model

Step-by-Step Methodology

Step	Description
Step 1	Load patient dataset and split into training (80%) and testing (20%)
Step 2	Preprocess structured and unstructured data
Step 3	Generate patient embedding vectors
Step 4	Apply similarity metric (Cosine Similarity)
Step 5	Retrieve Top-5 most similar historical cases
Step 6	Analyze treatment and outcomes of similar cases
Step 7	Predict probable treatment success rate
Step 8	Display similar case profiles with risk score

IV. CONCLUTION

This study proposes an AI-driven Patient Case Similarity System to support treatment decision-making for mucormycosis and cancer. The system integrates structured clinical data, imaging features, and similarity-based retrieval mechanisms to enhance personalized medicine [1], [2].

By leveraging nearest neighbor approaches [16] and deep learning-based feature representation [20], the proposed framework improves case comparison accuracy beyond traditional classification systems. Cancer analytics research has demonstrated the effectiveness of AI in diagnosis and prognosis [7], [9], while fungal infection studies highlight the importance of early clinical intervention [11], [12].

The integration of these approaches into a unified similarity-based architecture contributes to intelligent clinical decision support systems [24]. Future work will focus on incorporating explainable AI models and large-scale validation using multi-institutional datasets.

However, certain limitations remain, including data availability constraints, class imbalance issues, and the need for large-scale validated datasets. Future work will focus on incorporating deep learning-based patient embeddings, explainable AI mechanisms for improved interpretability, and real-time deployment within hospital Electronic Health Record systems.

V. ACKNOWLEDGMENT

The author would like to sincerely thank the Babu Banarasi das Institute of Technology, Lucknow, and the Department of Computer Science for providing the academic support, technical resources, and guidance necessary to complete this project. The author is also grateful to the professors and colleagues whose insightful comments, encouragement, and constructive criticism greatly enhanced the quality and applicability of this research. Their collaboration was invaluable in successfully integrating machine learning techniques with practical applications. Special thanks are extended to Mr Shekhar srivastav, Assistant Professor, Department of Computer Science, for her expert guidance, continuous support, and assistance throughout the course of this research project

REFERENCES

[1] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.
[2] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.
[3] A. Rajkomar, E. Oren, K. Chen et al., “Scalable and accurate deep learning for electronic health records,” npj Digital Medicine, vol. 1, no. 1, pp. 1–10, 2018.
[4] P. Weng, Y. Chen, and C. Li, “Patient similarity networks for precision medicine,” Journal of Biomedical Informatics, vol. 85, pp. 67–77, 2018.
[5] H. Avati, K. Jung, and N. Shah, “Improving palliative care with deep learning,” BMC Medical Informatics and Decision Making, vol. 18, no. 4, pp. 1–12, 2018.
[6] S. Sohn, J. Kim, and L. Li, “Comparison of distance metrics for clinical similarity search,” IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1285–1293, 2018.
[7] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115–118, 2017.
[8] J. Litjens et al., “A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.
[9] T. Albahri et al., “Systematic review of artificial intelligence techniques in cancer detection and diagnosis,” Computers in Biology and Medicine, vol. 122, 2020.

[10] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.
[11] R. Garg et al., “Coronavirus disease-associated mucormycosis: a systematic review,” Mycoses, vol. 64, no. 12, pp. 1452–1459, 2021
[12] A. Skiada et al., “Epidemiology and diagnosis of mucormycosis: an update,” Journal of Fungi, vol. 6, no. 4, 2020.
[13] N. Petrikkos et al., “Epidemiology and clinical manifestations of mucormycosis,” Clinical Infectious Diseases, vol. 54, pp. S23–S34, 2012.
[14] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient over F1 score,” BMC Genomics, vol. 21, no. 6, 2020.
[15] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[16] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
[17] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, 2019.
[18] O. Chapelle, B. Schölkopf, and A. Zien, Semi-Supervised Learning. MIT Press, 2006.
[19] F. Wang et al., “Clinical information retrieval using patient similarity,” AMIA Annual Symposium Proceedings, pp. 1926–1935, 2015.
[20] M. Shickel, P. Tighe, A. Bihorac, and P. Rashidi, “Deep EHR: a survey of recent advances,” IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589–1604, 2018.
[21] C. Shortliffe and J. Cimino, Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Springer, 2014.
[22] World Health Organization, “Cancer fact sheet,” WHO, 2023.
[23] American Cancer Society, “Cancer treatment and survivorship statistics,” 2022.
[24] R. K. Singh et al., “AI-based clinical decision support systems in healthcare,” IEEE Access, vol. 8, pp. 184–205, 2020
[25] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge University Press, 2009.