AMRITA VISHWA VIDYAPEETHAM

DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF THE UGC ACT, 1956

AMRITA SCHOOL OF ARTIFICIAL ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

April - 2025

B. TECH ARTIFICIAL INTELLIGENCE IN DATA SCIENCE AND MEDICAL ENGINEERING

PREDICTION OF RHEUMATOID ARTHRITIS SEVERITY USING BIOMARKERS AND BLOCKCHAIN

24AIM112 - Molecular biology & basic cellular physiology

24AIM115 - Ethics, innovative research, businesses & IPR

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE – 641 112

BONAFIDE CERTIFICATE

This is to certify that the report entitled **"Rheumatoid Arthritis Severity using Biomarkers and Blockchain"** submitted by:

Battari Pavani Shreeya	**CB.AI.U4AIM24106**
Kothavari Nitya	**CB.AI.U4AIM24123**
Ardhra Vinod	**CB.AI.U4AIM24105**
Hari Krishna Iyer	**CB.AI.U4AIM24114**

for the final project of 2nd semester in **B. Tech Artificial Intelligence and Data Science in Medical Engineering** is a Bonafide record of the work carried out at Amrita School of Artificial Intelligence, Coimbatore.

Submitted for the final evaluation on 23/4/2025

Faculty **Faculty**

Project Report

RHEUMATOID ARTHRITIS SEVERITY USING BIOMARKERS AND BLOCKCHAIN

ABSTRACT:

Rheumatoid Arthritis (RA) is a chronic autoimmune disorder involving ongoing inflammation, which can lead to joint damage and disability. Predicting RA severity according to cytokine biomarkers allows clinicians to make timely and tailored treatment decisions. This work presents a GUI-based RA severity predicting system based on an XG-Boost machine learning model trained from clinically meaningful serum biomarkers such as IL-6, IL-17, TNF-Alpha, IL-10, and CD4+ levels. The model predicts RA severity into Mild, Moderate, or Severe categories, along with medication suggestions. Furthermore, to maintain data integrity, confidentiality, and traceability, a streamlined blockchain architecture is implemented to securely store prediction outcomes and medical choices. The system includes user authentication for authorized use, a clear graphical user interface (GUI) for data entry and visualization, and a persistent history of past predictions. This interface is designed to assist clinicians in more accurate, data-based decision-making in the treatment of RA.

INTRODUCTION:

Rheumatoid arthritis is an autoimmune disease that affects one percent of the world population is affected, leading to progressive destruction of joints, such as impairments of function and systemic complications. To achieve patient-based outcomes in therapy as well as preventing irreversible damage, the assessment of activity and severity in RA is important.

Moreover, serum biomarkers such as IL6, IL17, and TNF-alpha, IL10, and CD4+ cell counts are central to the understanding of the inflammation and immunological activity happening in RA patients. The disease progression is assessed by the analysis of the particular biomarkers, which leads to the modification of treatment strategies.

This project developed a machine learning based severity predicting system using XG-Boost, a robust and scalable gradient boosting framework. The training of the model was done by using a database of biomarkers observed in serum and their resulting categorized severity for RA, which yielded highly predictive rates. To enhance clinical practice, a GUI was developed to receive patient profiles on biomarkers in order to provide severity prediction and medication suggestions.

Addressing the critical need for security, integrity, and traceability in medical applications, a simple blockchain system was then incorporated. The blockchain will securely store the patient's predicted data and treatment.

The system significantly improves the speed and reliability of assessment of RA severity and provides a framework for implementing relatively secure practices in data handling in health care environments. It forms a decision-support tool meant to aid rheumatologists delivering personalized, evidence-based care.

LITERATURE REVIEW:

Topic	Study Title & Reference	Methods	Key Findings	Limitations
Diagnostic Biomarkers for RA	Identification of diagnostic biomarkers of rheumatoid arthritis based on machine learning-assisted comprehensive bioinformatics and its correlation with immune cells (PubMed: https://pubmed.ncbi.nlm.nih.gov/39170142/)	Utilized LASSO, SVM-RFE, and Random Forest algorithms on gene expression datasets; validated findings using ROC analysis.	Identified seven potential RA biomarkers (KYNU, EVI2A, CD52, C1QB, BATF, AIM2, NDC80) with AUC > 0.8; found correlations with immune cell infiltration patterns.	Limited by dataset size, lack of external validation, and absence of detailed clinical information.
Machine Learning in RA Management	Machine learning approaches to improve disease management of patients with rheumatoid arthritis: review and future directions (PubMed: https://pubmed.ncbi.nlm.nih.gov/34890271/)	Reviewed ML applications in RA diagnosis, monitoring, and treatment prediction; analyzed various ML models including Random Forests and SVMs.	ML shows promise in enhancing RA management through improved diagnostics and personalized treatment plans.	Challenges include data standardization, overfitting, ethical concerns, and limited real-world application.
Protecting Patient Data with Blockchain	How Healthcare Is Applying Blockchain Technology (Investopedia: https://www.investopedia.com/tech/how-health-care-moving-toward-blockchain/)	Discussed blockchain's role in securing health information exchanges and improving data interoperability.	Blockchain offers secure, immutable, and transparent data management, potentially unifying patient histories and maintaining strict privacy controls.	Adoption faces obstacles like regulatory compliance, scalability issues, and integration with existing systems.
Ethical Landscape of Digital Biomarkers	Mapping the ethical landscape of digital biomarkers: A scoping review (PMC: https://pmc.ncbi.nlm.nih.gov/articles/PMC11098308/)	Conducted a scoping review to identify ethical issues in digital biomarker development and implementation.	Highlighted concerns about privacy, informed consent, data ownership, algorithmic transparency, and regulatory gaps.	Noted a lack of standardized ethical guidelines and the need for more focused ethical analyses in this emerging field.

METHODOLOGY:

1. Data Preparation & Label Creation

- **Biomarker Conversion:** Raw biomarker values are taken as input and then converted to standardized serum levels using specific formulas:
    - IL-6: 5 × raw_value
    - IL-17: (2^log2_value) × 1.2
    - TNF-α: (2^log2_value) × 1.5
    - IL-10: 3.5 × rma_value
- Labeling for severity prediction is done as follows:
    - The system predicts that the person has Severe RA if TNF-alpha is greater than 10 and IL-6 is greater than 40 or IL-10 is greater than 25 and IL-17 is greater than 20 or CD+4 is greater than 6.
    - The system predicts that the person has Moderate RA if TNF-alpha is greater than 5 and IL-6 is greater than 10 or IL-10 I greater than 10 and IL-17 is greater than 8 or CD+4 is greater than 3.
    - If the above two statements are not satisfied then the system predicts that the person has Mild RA.

2. Feature Engineering

- The selected biomarkers are IL-6, IL-17, TNF-alpha, IL-10 and CD+4 are selected features.
- Severity labels are Severe, Moderate, Mild.

3. Data Preprocessing

- Text labels are converted to numerical values (Severe=0, Moderate=1, Mild=2)

- Using Min Max Scaler raw biomarker values are normalized to a range of [0,1].

4. Data Splitting

For Training and Testing the data is split into 80-20.

5. Model Training

- Using XG-Boost classifier with optimized parameters the training is done.
- Training Process is using early stop based on validation loss, the loss curve monitoring, and accuracy evaluation on test set is done.

6. Model Deployment

The trained model is saved in the operating device as xgboost_RA_severity_model.json file, the Feature scaler is saved as scaler.pk1 and the Label encounter as label_encoder.pk1.

7. Inference Pipeline

- The raw inputs are converted using clinical formulas.
- Saved scaler file is used to scale features.
- Decode prediction to text label.

The pipeline demonstrates a robust implementation of machine learning for clinical decision support, following best practises in data preprocessing, model evaluation, and deployment reliability. The tight integration between domain knowledge(biomarker conversion rules) and data-driven modelling enables clinically interpretable predictions.

BUILDING GUI:

- Using Tkinter in python the Graphical User Interface (GUI) is built for entering biomarker levels(5) and get prediction using the trained model.
- Tkinter is a standard GUI library for Python which provides a fast and easy way to create desktop applications.
- The pre-trained model was loaded for making predictions.
- The application window was given a sky blue background.
- 5 Input fields were created and a “prediction severity” button was added to get prediction.

- This GUI takes input biomarker levels, scales them, makes a prediction and displays the result.
- This GUI is useful for lab researchers, clinicians and doctors to analyze patients with RA severity.

BLOCKCHAIN INTEGRATION:

- **Data Bundling**
  - Each time a user inputs biomarker levels and gets a severity prediction, the data is bundled into a "block" along with a timestamp.
- **Blockchain Structure**
  - These blocks are added to a secure, linked digital ledger (the "blockchain").

  Immutability
  - Once added, the blocks cannot be changed, ensuring tamper-proof data.
- **Data Integrity**
  - This prevents alteration of past medical records, ensuring accurate and trustworthy information.
- **Transparency & Security**
  - Blockchain guarantees full transparency, security, and integrity, critical for sensitive health data tracking.

Libraries used in python:

pandas for data handling

numpy for math operations

xgboost for machine learning

jonlib for saving/loading models

hashlib for blockchain security

tkinter for creating the user interface

ETHICAL STANDARDS:

The ethical standards taken into consideration when dealing with AI based systems are:

a. The system is designed with minimum complexity and more user friendliness so that can serve the purpose respecting human rights and enhancing their capabilities it by serving their needs.
b. User trust is gained by providing transparency ensuring of how the system takes decisions. The model counts on moral human society and human rights.
c. Given to its transparency user might think that the system is not safe for use but our first intention is to be private. Our system is Tamper-evident which is by the blockchain integration.

i. ***Problems with Ethical Approval and How to Fix Them: Lessons from Three Trials in Rheumatoid Arthritis***

Authors: Jonathan Mendel, Ben Goldacre, Edzard Ernst and Samuel Whittle

This research paper has a focus on gaps, in other words, ethical oversight and transparency in clinical trials. The two trials under consideration are those of ocrelizumab (namely STAGE and FEATURE), and the last one is dedicated to homeopathy. There were barriers to gaining access to documents from ethics, and at first, there was reluctance to disclose files entirely or for the part of such files by pharmaceutical companies and NHS trusts: in addition, usage of placebo despite availability would be criticized, as those included were at risk of being damaged irreversibly. Homeopathy trial was equally worrisome for not meeting the standards where inclusion/exclusion criteria were concerned, along with misleading hints at the remedy composition.

The committee approved so that the trials were even looming without securing proper informed consent in most cases, especially when the risks of placebo treatments did not clearly reflect their methodological limits. The authors argue that such an endorsement should not be assumed to be conclusive evidence of ethical soundness: they would make reforms on sharing ethics documents transparently, provision of summaries of preexisting evidence and trial quality to

participants, risk mitigation (especially with placebo), and accountability through open access to consent forms and committee correspondence.

Ethical Concerns discussed in this paper

However, certain ethical implications encumber the work. Most particularly, knowing that methotrexate would not work after a prior failure in some, but withholding appropriate alternatives such as rituximab with no exceptions allowed rescue therapy, members of the placebo finance arm received very little rescue therapy. However, the informed consent was supposed to advise the subject in advance of additional risks associated with placebo treatment such as pain and permanent joint damage.

There were inconsistencies between the terms of the ethics committee forms and that of the research protocol for the homeopathic trial leading to vague eligibility criteria while giving the impression that homeopathic remedies would be 'active ingredient-infused' misrepresentation of the procedure. All in all, the article shows how potential ethical considerations of a patient review miss and how patients tend not to receive integral and reliable information. The authors further propose more transparency and demanding scrutiny to guarantee patients' safety and trust in clinical research.

ii. ***Artificial Intelligence in Rheumatoid Arthritis: Potential Applications and Future Implications***

Authors: Vinit J. Gilvaz1 * and Anthony M. Reginato1,2

This paper describes how AI-and ML and DL-more specifically are impacting RA management. Diagnosis, flare prediction, disease activity monitoring, and treatment decision-making are areas where AI models are put to work. Neural networks and CNNs may operate on heterogeneous data types consisting of clinical, genetic, and imaging data to totally automate the diagnosis (e.g., hand radiographs), identification of RA subtypes, and even detection of joint damage. AI aids in predicting treatment response and patient stratification in hard-to-treat cases of RA. Real-time flare detection and clinical outcome prediction will find more clinical application through smart integrations of wearable devices and

EHR-integrated models. The capability of AI may, in the future, also allow this technology a real chance in RA personalized care for diminishing treatment trial-and-error and solving diagnosis.

Ethical Concerns discussed in this paper

In this paper, the authors pointed out various ethical and practical challenges related to the incorporation of AI in the care of RA patients. First, data privacy raises major hurdles because patient health records are confidential, making it hard to use these for training of models. Therefore, biased and unstructured data about healthcare may cast doubt over predictions' reliability. Thirdly, it is risky to use AI models, most notably deep learning, since they tend to be black boxes that give little explanation of their decisions in an area, namely medicine, where transparency is critical. The authors further point out gaps in regulations, saying that the FDA and the European Union have commenced efforts to formulate guidelines, for example, the FDA's AI/ML SaMD Action Plan, but regulation still remains in the developmental stage. Finally, one point stressed in the paper is the need for doctors to be involved in, and educated about the use of, AI to allow for safe, interpretable, and ethically accountable AI integration into standard clinical care.

iii. ***On the Integration of Artificial Intelligence and Blockchain Technology: A Perspective About Security***

Authors: Oleksandr Kuznetsov, Paolo Sernani, Luca Romeo, Emanuele Frontoni, Adriano Mancini

The article entitled "On the Integration of Artificial Intelligence and Blockchain Technology: A Perspective About Security" explores the potential synergy of Artificial Intelligence (AI) and Blockchain technology and how their collaborative use can exponentially increase system security, reliability, and performance. AI is capable of analyzing large volumes of data, identifying patterns, and making independent decisions, while blockchain provides a decentralized, unalterable, and open environment for storing and managing data. Their fusion enables groundbreaking applications in an enormous array of fields like healthcare, finance, supply chains, cybersecurity, and smart cities. For example, blockchain can give a reliable environment to ensure AI decision verification, hence fixing

the issue of transparency, whereas AI can optimize the performance of blockchain by way of smart optimization and analytics. The paper, however, also recognizes some technical challenges to the technologies such as scalability limitations, interoperability issues across platforms, and the computational cost that comes with marrying these technologies together. It is possible that deliberate architectural design, continuous research, and the evolution of hybrid systems could be a precursor to fruitful implementation of AI-blockchain technology in practical applications.

Ethical Concerns discussed in this paper

The paper also raises a nuanced discussion on the ethical ramifications of combining AI and blockchain technologies. One of the major issues has to do with privacy and information protection. Although blockchain's immutability prevents data from being tampered with, it also means that when sensitive or personal data is kept on-chain, it is there forever and available, which undermines privacy rights like the "right to be forgotten." In addition, the combination of AI might enhance current ethical issues, including bias in algorithmic decision-making, particularly where those decisions are made on a blockchain and form part of an immutable record. The authors stress that ethical AI needs to be transparent, explainable, and programmed to avoid discrimination or bias. They also reinforce accountability frameworks and regulatory frameworks to prevent misuse of these technologies. Lacking governance, the synergistic effect of AI and blockchain will unknowingly result in enhanced surveillance, loss of control, or abuse in high-stakes applications such as healthcare diagnosis or financial choices. Thus, the paper advocates for an integrated and morally sound approach to the design and deployment of AI-blockchain systems.

iv. ***Machine Learning-Based Remission Prediction in Rheumatoid Arthritis Patients Treated with Biologic Disease-Modifying Anti-Rheumatic Drugs: Findings from the Kuwait Rheumatic Disease Registry***

Authors: Ahmad R. Alsaber, A. Al-Herz, B. Alawadhi, I. A. Doush, P. Setiya, A. T. AL-Sultan, K. Saleh, A. Al-Awadhi, E. Hasan, W. Al-Kandari, K. Mokaddem, A. A. Ghanem, Y. Attia, M. Hussain, N. AlHadhood, Y. Ali, H. Tarakmeh, G. Aldabie, A. AlKadi, H. Alhajeri

This study focuses on leveraging **machine learning (ML)** models to predict **remission outcomes** in **rheumatoid arthritis (RA)** patients undergoing treatment

with **biologic DMARDs (bDMARDs)**, based on data from the Kuwait Registry for Rheumatic Diseases (KRRD).

ML models used included:

- Lasso
- Ridge regression
- Support Vector Machine (SVM)
- Random Forest
- XGBoost

They also implemented **Shapley Additive Explanations (SHAP)** to identify key predictors influencing remission across various bDMARD treatments.

Data was collected from **1,968 patients** treated between 2013 and 2022 across four public hospitals in Kuwait.

Ethical Concerns Discussed in This Paper:

Although the paper does not explicitly focus on ethical shortcomings, it presents several implicit ethical considerations that are critical in the context of machine learning applications in healthcare. One major concern is **data privacy**, as handling sensitive clinical and patient information through a national registry like the Kuwait Registry for Rheumatic Diseases requires strict adherence to consent protocols and secure data management practices. Another important ethical aspect is **model transparency**—the study's use of explainable artificial intelligence tools, such as SHAP (Shapley Additive Explanations), helps ensure that the predictions made by machine learning models are interpretable and justifiable in clinical settings. This supports the ethical obligation to provide clear, understandable insights to healthcare providers and potentially to patients. Additionally, the impact on **patient outcomes** must be considered, as predictive models can significantly influence therapeutic decisions. Inaccurate predictions could lead to inappropriate treatments or delays in effective care, thereby affecting the patient's well-being. Through this study, the authors underscore the importance of ethically grounded, data-driven methodologies to enhance both decision-making processes and the overall efficiency of healthcare systems.

V. ***Transforming Rheumatoid Arthritis Management: Harnessing Artificial Intelligence for Early Detection, Personalized Treatment, and Ethical Challenges***

Authors: Priyabrata Thatoi, Rohit Choudhary, Sushree Swapnil Rout

This review paper discusses about the role of artificial intelligence (AI) in improving rheumatoid arthritis (RA) management, focusing on early detection, personalized treatment, and ethical challenges. AI techniques like machine learning (ML), deep learning (DL), and natural language processing (NLP) are shown to improve diagnostic accuracy by using multi-omics data (genomic, proteomic, metabolomic) and imaging analyses. ML models such as Elastic Net and Random Forest shows high predictive accuracy (AUC up to 0.97) in identifying RA biomarkers and disease progression. Imaging tools which involve convolutional neural networks (CNNs), automate joint erosion detection in X-rays and MRIs, matching expert rheumatologist performance. AI also enables personalized treatment by predicting drug responses and supports drug discovery by accelerating target identification and repurposing existing therapies. These advancements highlight AI's potential to shift RA care from reactive symptom management to proactive, precision medicine.

The paper also discusses about ethical and technical challenges in AI deployment. Biases in training data, lack of model transparency, and privacy concerns are the potential risks. Solutions include fairness-aware algorithms and interpretability frameworks are critical to ensuring equitable outcomes. AI-driven tools for remote monitoring (wearables, telemedicine) and patient education (chatbots, self-management apps) further democratize access to care but require robust data governance. The review concludes that while AI significantly improves RA diagnosis, treatment personalization, and drug development, addressing ethical issues and fostering interdisciplinary collaboration are essential to fully realize its benefits in clinical practice.

Ethical Concerns discussed in this paper

The paper discusses about several important ethical issues in the use of AI for rheumatoid arthritis (RA) care, such as AI model bias that can lead to healthcare inequalities by being biased toward some populations (e.g., race, gender, socioeconomic status) in terms of diagnosis or treatment suggestions. Transparency and explainability are two essential challenges as they can hide decision-making, which can break clinician and patient trust. Privacy threats emerge with the management of sensitive patient information, especially in telemedicine and remote monitoring, requiring strong data governance to maintain confidentiality. To overcome these concerns, the authors suggest implementing fairness-aware algorithms to counteract biases, crafting interpretability frameworks for transparency, and implementing ethical standards for safe, fair AI implementation in healthcare.

vi. An overview of a recent court challenge to the protection of biomarkers as intellectual property

Authors: Stephen C. Hall, Justin M. Tromp, Saeed A. Jortani

This paper discusses about the legal and ethical implications of a landmark U.S. court case (Association for Molecular Pathology v. Myriad Genetics) that invalidated patents on the BRCA1 and BRCA2 genes, which are connected to hereditary breast and ovarian cancers. Myriad Genetics claimed patents on isolated DNA sequences and diagnostic methods for detecting mutations in these genes. The court told that naturally occurring DNA sequences, even when isolated, are unpatentable "products of nature" under Section 101 of the U.S. Patent Act. It also rejected method claims for genetic testing as abstract mental processes lacking transformative steps. The decision underscored that mere analysis or comparison of genetic information, without inventive application or physical transformation, to meet patent eligibility criteria.

It emphasizes that patent claims for biomarkers or genetic tests must include tangible, transformative to satisfy legal requirements. Practical advice for inventors includes drafting claims that integrate actionable steps beyond data analysis and aligning with precedents that emphasize practical utility. The case serves as a cautionary guide for scientists and patent counsel, stressing the need to balance legal strategies with ethical considerations and public access to genetic advancements.

Ethical Concerns discussed in this paper

The paper discusses ethical concerns surrounding the patenting of genetic biomarkers, particularly focusing on the monopolization of genetic information and its impact on healthcare access and scientific progress. Myriad Genetics' patents on BRCA1/BRCA2 genes raised issues of restricted patient access to affordable genetic testing, as the company controlled all diagnostic use, limiting alternatives. Critics argued this stifled research by preventing other scientists from studying or developing tests for these genes, undermining collaborative innovation. Also, the case highlighted ethical confusions over owning human DNA, as patents on naturally occurring sequences commodify biological material, conflicting with principles of shared human heritage. The court's invalidation of these patents aimed to prioritize public good over corporate control, ensuring broader access to genetic knowledge and fostering equitable advancements in personalized medicine.

RESULTS:

The trained XGBoost model was efficient enough in predicting RA severity with accuracy of 99.75%. The system could classify patients into mild, moderate and severe RA categories based on the threshold values of biomarkers.

The Tkinter-based GUI was successfully developed, allowing users to input biomarker values and get instant predictions.

Fig1 Login page for user

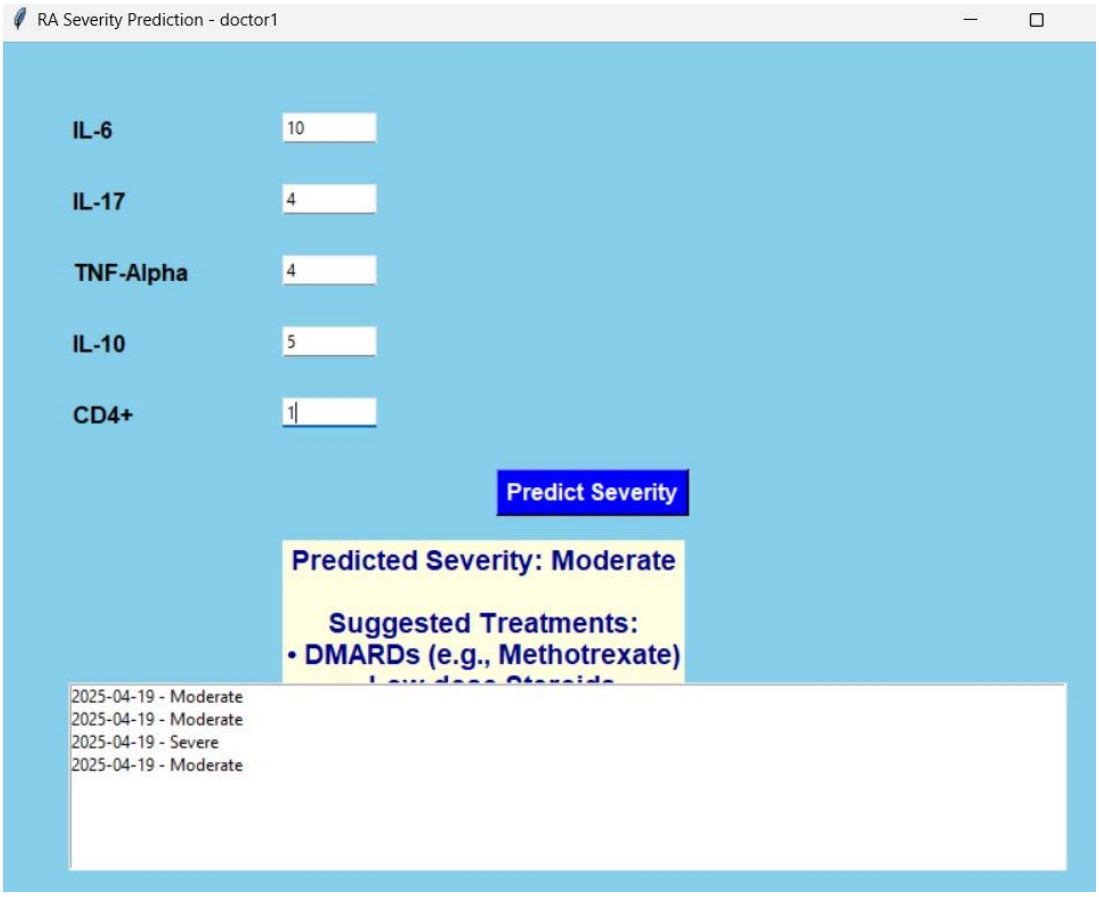
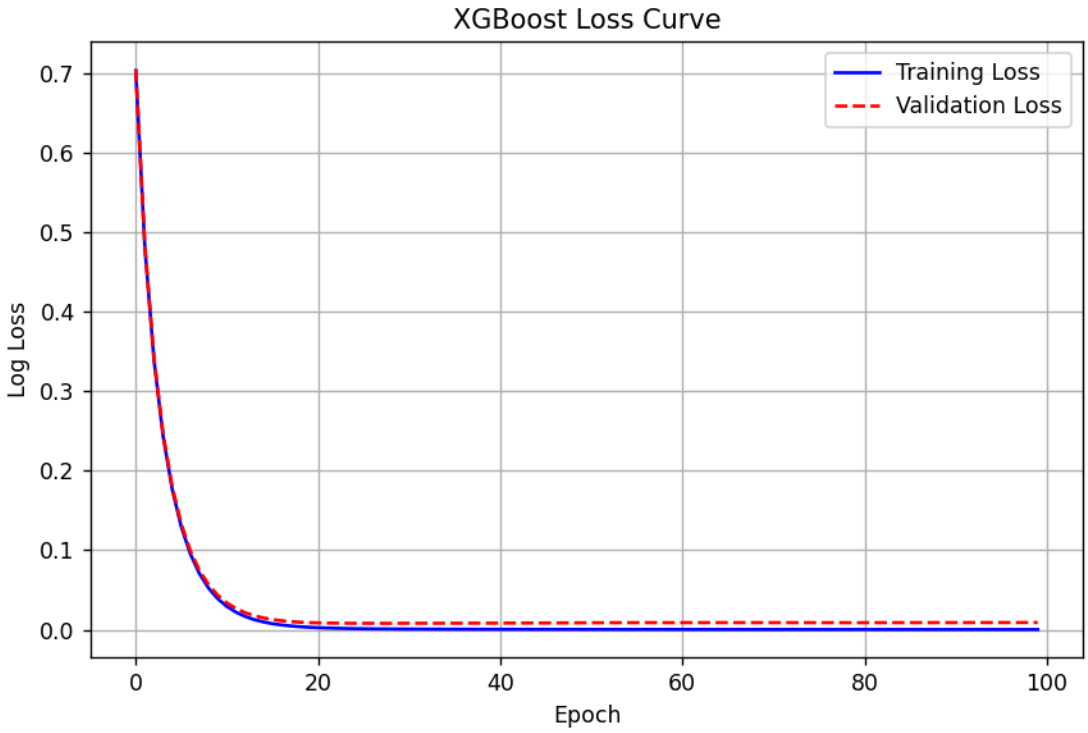


Fig2 Graphical User Interface (GUI)

Predicted results were then followed by specific medication suggestions, according to the severity of RA. Also patient data was securely stored using blockchain, ensuring privacy, immutability, and transparency.



CONCLUSION:

The project shows that using machine learning (XGBoost) with blockchain technology can improve RA diagnosis and management. The high prediction accuracy shows the reliability of biomarker-based AI models in clinical decision-making. The GUI improves accessibility, allowing clinicians and researchers to use the model in real-time. Blockchain integration ensures secure and tamper-proof medical records, improving trust in digital healthcare systems. So, this approach provides a cost-effective, accurate, and ethical method for RA severity prediction and treatment guidance.

REFERENCES

https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.963540/full

https://www.cell.com/heliyon/fulltext/S2405-8440(24)11542-3

https://www.tandfonline.com/doi/abs/10.1080/1744666X.2022.2017773

https://journals.lww.com/jehp/fulltext/2024/03280/securing_patient_data_in_the_healthcare_indu

stry_.94.aspx?context=latestarticles

https://www.sciencedirect.com/science/article/pii/S1570963914000028

https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000519