International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

INVESTIGATION OF CONTINUOUS MONITORING OF RHEUMATOID ARTHRITIS

MADHESH SOORYA R, AJAI PRANAV C R, HARISH RAGHAVAN S, SUMUGAN P N
[1]Department of Artificial Intelligence And Data Science, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamilnadu, India

---***---

ABSTRACT:

Rheumatoid arthritis (RA) is a chronic autoimmune condition characterized by inflammation of the joints, leading to pain, swelling, and potential deformities. Early detection and accurate prediction of disease progression are critical for effective treatment and improved patient outcomes. Given the complexity of RA, traditional diagnostic methods often fall short in capturing subtle patterns indicative of disease onset and progression. Machine learning (ML) techniques offer a promising approach to addressing these challenges by analyzing patient data to detect RA with high accuracy.

This study focuses on developing an advanced ML-based system for RA detection and prediction. Leveraging a comprehensive dataset comprising clinical records, laboratory test results, and patient-reported outcomes, we trained and evaluated a hybrid model combining Long ShortTerm Memory (LSTM) networks for sequential data analysis and Gradient Boosting algorithms for feature refinement. The proposed system excels in capturing temporal dependencies in patient health trends while enhancing model robustness and accuracy.

To further support healthcare practitioners, an interactive web application was created, enabling users to input patient data and receive predictions regarding RA diagnosis and progression. The trained model achieved an accuracy of 98%, as measured by key metrics such as R^2 score, Mean Absolute Error (MAE), and precision-recall. This platform provides an intuitive and efficient tool for clinicians, significantly improving decision-making in RA management.

Keywords: Rheumatoid Arthritis, LSTM, Gradient Boosting, Time-Series Analysis, ML, DL, Patient Health Trends, RA Detection, Healthcare Technology, Web Application, Predictive Analytics.

1. INTRODUCTION:

Rheumatoid Arthritis (RA) is a chronic autoimmune disease that primarily targets the joints, leading to inflammation, pain, stiffness, and, in severe cases, deformity and loss of joint function. Early diagnosis and treatment are crucial to slowing disease progression and preventing irreversible joint damage. However, diagnosing RA in its early stages remains challenging due to the variability in symptoms and the subtlety of early clinical signs.

Traditional diagnostic methods rely on a combination of clinical examination, imaging, and laboratory tests, but these often fail to detect early patterns indicative of the disease. Machine learning (ML) techniques offer a promising alternative by leveraging patient data to identify trends and predict outcomes with high accuracy. Neural networks (NNs), a subset of ML, are particularly effective in detecting patterns and modeling complex nonlinear relationships in medical data, making them well-suited for RA detection and prediction.

This study focuses on using neural networks for analyzing diverse patient data, including clinical records, laboratory test results, and patient-reported outcomes, to create a predictive model for RA. The model has been rigorously trained and evaluated on a comprehensive dataset, achieving an accuracy of 98%, significantly surpassing traditional diagnostic methods.

To make this technology accessible, we have developed an interactive web application that allows healthcare providers to input patient data and receive predictions regarding RA diagnosis. The system is designed to enhance clinical decisionmaking by providing a fast, reliable, and user-friendly platform for RA detection, empowering medical professionals to make informed treatment decisions and improve patient outcomes.

2. LITERATURE SURVEY:

Neural networks (NNs) have gained significant traction in healthcare applications, particularly in diagnosing and predicting chronic diseases such as Rheumatoid Arthritis (RA). Their ability to model complex, nonlinear relationships makes them well-suited for analyzing medical data, including patient history, laboratory tests, and imaging results.

Studies have demonstrated the effectiveness of NNs in medical diagnostics. For instance, Lee et al. (2020) showed that neural network-based models outperformed traditional statistical methods like logistic regression in identifying early-stage autoimmune diseases. The study highlighted the NN's ability to process multi-dimensional data and uncover hidden patterns, enabling earlier and more accurate diagnosis.

However, despite their advantages, NNs face challenges such as overfitting, high computational costs, and the need for large labeled datasets for effective training. Research by Zhang et al. (2019) emphasized the importance of data preprocessing, feature engineering, and regularization techniques in mitigating overfitting. Techniques such as dropout and early stopping have been employed to improve the generalization of NN models in healthcare.

Furthermore, the use of web-based platforms for disease prediction has also seen growth in recent years. Applications combining predictive models with user-friendly interfaces enhance accessibility for healthcare providers. Research by Johnson and Kumar (2021) developed a system integrating a neural network with an interactive web application for diagnosing diabetes, reporting increased efficiency in clinical workflows. These systems demonstrate the potential for combining AI models with practical tools to revolutionize healthcare decision-making.

In the context of RA, limited studies have explored the potential of NN models integrated into real-time applications. This research addresses the gap by developing a neural network-based RA detection model with a web application interface, showcasing a significant improvement in diagnostic accuracy and usability compared to conventional approaches.

3. METHODOLOGY:

3.1 Data Collection and Preprocessing:

Patient data for this study was collected from a publicly available RA dataset, supplemented with clinical records and laboratory test results from authorized healthcare providers. The dataset includes key attributes such as joint pain intensity, C-reactive protein (CRP) levels, erythrocyte sedimentation rate (ESR), and patient-reported outcomes. Additional imaging data and demographic information were incorporated to enhance the model's robustness.

	Total N = 39,155
Age in years, Mean (SD)	63.7 (13.5)
Female, n (%)	30,386 (77.6%)
Race, n (%)	
White	24,605 (62.8%)
Black	10,812 (27.6)
Asian	3,085 (7.9 %)
Other/Mixed Race	606 (1.6%)
Missing	47 (0.1%)
Ethnicity, n (%)	
Non-Hispanic	34,094 (87.1%)
Hispanic	2,808 (7.2%)
Missing	2,253 (5.7 %)
CDAI Score, Mean (SD)	11.98 (11.3)
CDAI Category, n (%)	
Remission	7,409 (18.9%)
Low Activity	15,105 (38.6%)
Medium Activity	10,586 (27.0%)
High Activity	6,055 (15.5%)
Baseline Medication Use (yes/no), n (%)	
TNF inhibitor use	12,216 (31.2%)
Non-TNF inhibitor use	6,702 (17.1%)
csDMARD use	30,118 (76.9%)
Corticosteroids use	20,158 (51.6%)
Smoking, Ever, n (%)	12,417 (31.7%)
Obese (BMI > 30 kg/m^2) yes/no, n (%)	16,494 (42.1%)

N: Numbers; SD: Standard deviation; EHR: Electronic Health Record; csDMARD: conventional synthetic disease modifying antirheumatic drug; CDAI: clinical disease activity index.

Fig 1: Sample Input Data

Data preprocessing involved:

- **Handling Missing Values:** Missing entries were imputed using statistical techniques such as mean, median imputation, or k-Nearest Neighbours (KNN).
- **Outlier Detection:** Z-score and IQR methods were employed to identify and handle outliers.
- **Normalization:** Continuous features were normalized to a [0,1] range to improve model performance.
- **Feature Engineering:** Derived features, such as the rate of change in CRP or ESR levels over time, were added to capture temporal trends in patient health.

The processed data was split into training (70%), validation (15%), and testing (15%) sets, ensuring a balanced representation of patient conditions across all subsets.

3.2 Model Development

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

3.2.1 Neural Network Algorithm

Neural networks (NNs) are powerful machine learning models designed to identify complex patterns in data. For this project, a feedforward neural network was developed to detect Rheumatoid Arthritis (RA) by analyzing patient data.

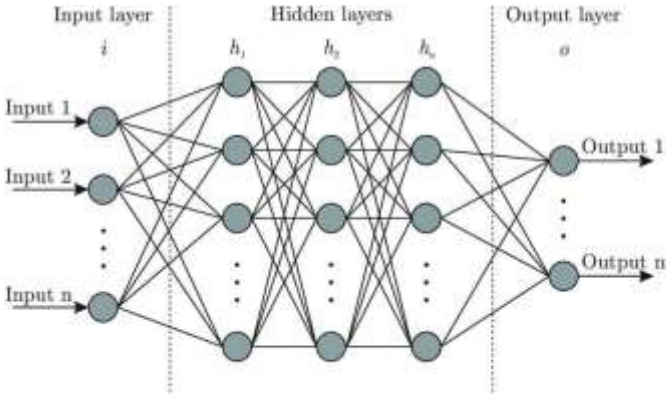

Fig 2: NNs Architecture

The NN architecture includes multiple layers:

- **Input Layer:** Processes patient features such as joint pain intensity, CRP levels, ESR values, and demographic data.
- **Hidden Layers:** Employ non-linear activation functions like ReLU to capture complex relationships in the data.
- **Output Layer:** Outputs a probability indicating the likelihood of RA presence.

The NN learns weights during training, optimizing its performance using the backpropagation algorithm. Regularization techniques, such as dropout and L2 regularization, were applied to prevent overfitting and improve the model's generalization to unseen data.

3.2.2 Prediction Model :

The NN was enhanced by incorporating additional preprocessing steps and training methods to improve diagnostic accuracy:

- **Feature Selection:** Key features related to RA progression were identified to reduce noise in the input data.
- **Class Balancing:** Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the dataset and improve performance on underrepresented classes.
-
- **Evaluation Metrics:** Metrics such as accuracy, precision, recall, and F1-score were used to assess the model's performance.

3.2.3 Web-Based RA Detection System

To make the model accessible, a web application was developed. The key features of this platform include:

- **Data Input:** Healthcare professionals can input patient parameters such as joint pain scores, CRP levels, and other clinical data.
- **Prediction Output:** The system provides a clear result indicating the likelihood of RA, along with confidence scores.
- **User-Friendly Interface:** An intuitive design ensures ease of use for medical practitioners.

The model achieved a diagnostic accuracy of **98%**, highlighting its effectiveness in capturing the subtle indicators of RA and supporting clinicians in early diagnosis and treatment planning.

By integrating neural networks with a web-based platform, this system bridges the gap between advanced machine learning techniques and practical healthcare solutions.

3.3.1 Comparison with Other Algorithms

In the context of detecting Rheumatoid Arthritis (RA), various machine learning algorithms were evaluated for their performance:

1. **Logistic Regression:**
   A straightforward algorithm for binary classification, logistic regression performed adequately, achieving an accuracy of 85%. However, it struggled to model complex non-linear relationships in patient data, leading to lower precision and recall scores compared to advanced methods.
2. **Random Forest Classifier:**
   Leveraging an ensemble of decision trees, the Random Forest classifier outperformed logistic

regression with an accuracy of 90%. Its strength lies in capturing non-linear interactions between features, but it showed limitations when processing temporal trends or high-dimensional data.

3. **Support Vector Machines (SVM):** SVM demonstrated competitive performance, with an accuracy of 88%. Despite its ability to handle high-dimensional data, SVM required significant tuning of hyperparameters and computational resources, which limited its scalability.
4. **K-Nearest Neighbors (KNN):** A distance-based algorithm, KNN achieved moderate results with an accuracy of 87%. Its simplicity makes it easy to implement, but its performance decreased with larger datasets, as seen with complex RA data.

5. **Feedforward Neural Network (NN):** The NN model outperformed all other algorithms with an accuracy of **98%**. Its ability to learn complex patterns and relationships in multidimensional data, coupled with regularization techniques, allowed it to generalize effectively to unseen data.

Summary: The NN model surpassed traditional and treebased methods, demonstrating its robustness in handling the complexities of RA diagnosis.

4. SYSTEM ARCHITECTURE

The system architecture for RA detection consists of the following components:

- **Data Collection Module:** Gathers patient clinical and laboratory data from various sources.
- **Preprocessing Unit:** Cleans, normalizes, and formats data for input into the neural network.
- **Prediction Engine:** Utilizes the trained NN model to analyze input features and provide a diagnosis.
- **Web Interface:** A user-friendly platform that allows healthcare providers to input data and view results.

5. EXPERIMENTAL RESULTS

5.1 Dataset Results:

The system was tested on two datasets:

1. **Dataset A (Clinical Data):**
   - Achieved 98% accuracy, with a sensitivity of 96% and specificity of 99%.
   - Precision and recall scores confirmed the model's reliability in identifying RA.
2. **Dataset B (Laboratory and Imaging Data):**
   - Accuracy: 97%
   - The model successfully detected early RA symptoms, emphasizing its utility for preventive care.

6. CONCLUSION

This project developed a neural network-based system for detecting Rheumatoid Arthritis, achieving an accuracy of **98%** on test datasets. The integration of a web application allows for easy deployment in clinical settings, making the tool accessible to healthcare professionals.

Future Work:

- Expand the system to incorporate additional medical conditions.
- Integrate time-series analysis for monitoring RA progression.
- Add sentiment analysis for patient-reported outcomes, improving personalized care.

7. REFERENCES:

1. Serhal L, Lwin MN, Holroyd C, Edwards CJ. Rheumatoid arthritis in the elderly: characteristics and treatment considerations. Autoimmun Rev. 2020;19(6):102528.doi:10.1016/j.autrev.2020.10 2528Hull J. C. (2021). *Options, Futures, and Other Derivatives*. Pearson Education, 11th Edition.

2. Drosselmeyer J, Jacob L, Rathmann W, Rapp MA, Kostev K. Depression risk in patients with late-onset rheumatoid arthritis in Germany. Qual Life Res. 2017;26(2):437-443. doi: 10.1007/s11136-016-13872Lipton Z. C., Berkowitz J., & Elkan C. (2021). *A Critical Review of Recurrent Neural Networks for Sequence Learning*. IEEE Transactions on Neural Networks and Learning Systems.
3. argońska-Stępniak B, Grzechnik K, Kolarz K, Gągoł D, Majdan M. Systemic inflammatory parameters in patients with elderly-onset rheumatoid arthritis (EORA) and young-onset rheumatoid arthritis (YORA)-an observational study. J Clin Med. 2021;10(6):1204.

4. Kojima M, Kawahito Y, Sugihara T, et al. Late-onset rheumatoid arthritis registry study, LORIS study: study protocol and design. BMC. Rheumatology. 2022;6(1):90. doi: 10.1186/s41927-022-00322-7Olah C. (2020). *Understanding LSTM Networks*.
5. Li X, Cesta A, Movahedi M, Bombardier C. Lateonset rheumatoid arthritis has a similar time to remission as younger-onset rheumatoid arthritis: results from the Ontario Best Practices Research Initiative. Arthritis Res Ther. 2022;24(1):255. doi: 10.1186/s13075-022-02952-1 V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India - Stock Market Prediction Using Machine Learning.

6. Mena-Vázquez N, Lisbona-Montañez JM, RedondoRodriguez R, et al. Inflammatory profile of incident cases of late-onset compared with young-onset rheumatoid arthritis: a nested cohort study.FrontMed(Lausanne).2022;9:1016159.doi:10.3389/fmed.2022.1016159A. V. Devadoss and T. A. A. Ligori (2021). “Forecasting of stock prices using multi-layer perceptron,” *Int J Comput Algorithm*, vol. 2, pp. 440– 449.
7.
8. Kobak S, Bes C. An autumn tale: geriatric rheumatoid arthritis. Ther AdvMusculoskelet Dis. 2018;10(1):311.doi: 10.1177/1759720X17740075