IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# CE-24 Multimodal AI-Based Arthritis Detection System Using Deep Learning

**Jui Ramteke[1], Achal Khobragade[2], Pranjali Chiwande[3], Pallavi Adbale[4], Pranoti Munjankar[5], Dr. Vanita Buradkar[6]**

[1]*Dr. Vanita Buradkar ::0000- 0001- 8033-4198*
*Rajiv Gandhi College of Engineering, Research & Technology, Chandrapur,Maharashtra, India, 442403*
[2]*Student, Jui Ramteke*
*Rajiv Gandhi College of Engineering, Research & Technology, Chandrapur,Maharashtra, India, 442403*

*Email of corresponding Author: juiramteke20@gmail.com*

***Abstract*** *– This paper presents a multimodal AI-based arthritis detection system integrating five independent diagnostic modules: a Vision Module (knee X-ray classification, spine MRI degeneration analysis, and synovial fluid microscopy), a Lab Report Module, a Wearable Sensor Module, a Genomics Module, and a Voice Emotion Detection Module. The Vision Module employs DenseNet121 trained on 5,778 knee X-ray images, achieving 82.40% classification accuracy across five Kellgren-Lawrence (KL) grades. Spine MRI degeneration is assessed via signal intensity analysis, while synovial fluid inflammation is detected using ResNet50 with K-Means clustering. The Lab Report Module applies Random Forest classification on clinical biomarkers. The Wearable Module employs a neural network on MotionSense sensor data for risk stratification. The Genomics Module classifies arthritis-associated SNP markers, and the Voice Module detects pain and fatigue using MFCC-based feature extraction. All modules produce structured outputs integrated into an automated PDF medical report generator, providing a scalable, interpretable, and accessible clinical decision support system.*

***Keywords-*** *Arthritis detection, deep learning, DenseNet121, multimodal AI, Kellgren-Lawrence grading, ResNet50, MFCC, genomics, wearable sensors, clinical decision support*

## INTRODUCTION

Arthritis refers to a group of joint diseases, over 100 different types, that are characterized by inflammation, stiffness, pain, swelling, and functional loss. Undiagnosed or inadequately treated arthritis leads to joint disability and a decrease in the patient's quality of life. There are three common forms: Rheumatoid Arthritis (RA) as an autoimmune disorder in which the immune system targets its own joints' synovial lining; Osteoarthritis (OA) as a progressive erosion of articular cartilage that leads to degenerative changes in joints; and Gouty Arthritis, which develops from an inflammatory response following the deposition of needle-like monosodium urate crystals within the joints. Prompt and accurate diagnosis is crucial in preventing further deterioration and maintaining musculoskeletal well-being.

Traditional arthritis diagnosis includes interpreting medical imaging studies, conducting clinical assessments, and ordering lab tests. All these methods suffer several major disadvantages: inconsistencies in clinician opinion, time-consuming procedures, absence of standard measures for quantification, and insufficient availability of specialists in rural areas. While current artificial intelligence applications demonstrate significant progress in assisting in diagnosing arthritis, they focus on one of the diagnostic modalities and do not integrate various sources of data available to clinicians, thus limiting their usefulness.

In order to overcome these problems, the proposed multimodal AI-based Arthritis Detection System incorporates five independent modules for diagnosing arthritis: (1) a Vision Module that consists of knee KL grading and spine MRI degeneration detection algorithms and the microscopic analysis of synovial fluids; (2) a Lab Report Module based on analyzing laboratory markers of disease; (3) a Wearable Sensor Module for assessing risk factors; (4) a Genomics Module for assessing SNP-based genetic risks;

and (5) a Voice Emotion Detection Module that detects signs of pain and fatigue. Automatic generation of PDF diagnostic reports allows clinicians to make decisions based on the obtained results.

## LITERATURE REVIEW

Substantial research has been conducted on AI-assisted arthritis detection, primarily using single-modality approaches. Bai et al. [1] developed an Artificial Neural Network (ANN) trained on six clinical biomarkers---age, sex, RF, anti-CCP, 14-3-3eta, and anti-CarP---achieving 90.6% accuracy for RA diagnosis. While this work demonstrates the utility of clinical data in AI-driven diagnosis, it is limited to tabular inputs and does not incorporate imaging or multimodal signals.

Wang et al. [2] introduced the Key-Exchange Convolutional Auto-Encoder (KECAE), a novel data augmentation strategy for early knee osteoarthritis detection. By separating and exchanging pathological features between X-ray images to synthesize clinically realistic training samples, the approach improved model accuracy by approximately +1.98% and was validated by radiologists.

Peng et al. [3] conducted a comparative evaluation of GoogLeNet, VGG16, and ResNet50 on hand radiographs for RA detection and severity staging, with GoogLeNet achieving 97.8% AUC and 100% sensitivity. However, multi-stage classification remained challenging.

Jose et al. [4] benchmarked CNN, SNN, SVM, VGG16, and Google Teachable Machine for KL-grade knee OA classification, reporting 77% accuracy for Grade 4 (severe) but significantly lower performance (30-40%) for early-stage grades due to class imbalance.

A critical gap across existing literature is the absence of a unified multimodal diagnostic framework. Prior systems focus on single imaging modalities, lack genomic and wearable integration, and provide limited explainability---factors that restrict clinical adoption. The proposed system addresses each of these limitations through modular multimodal architecture.

| S.No | Title | Authors | Methodology | Key Result |
|---|---|---|---|---|
| 1 | Improved Diagnosis of RA Using ANN | Bai et al. (2022) | ANN on 6 clinical biomarkers (age, sex, RF, anti-CCP, 14-3-3eta, anti-CarP) | 90.6% accuracy |
| 2 | KECAE for Early Knee OA Detection | Wang et al. (2025) | Feature-exchange augmentation on X-rays to generate synthetic images | +1.98% accuracy improvement |
| 3 | Radiograph-Based RA Diagnosis via CNN | Peng et al. (2024) | GoogLeNet, VGG16, ResNet50 on hand X-rays | GoogLeNet: 97.8% AUC, 100% sensitivity |
| 4 | ML-Based Diagnosis of Knee OA | Jose et al. (2025) | CNN, SVM, VGG16, Google Teachable Machine for KL grading | 77% for Grade 4; 30-40% for early grades |

*TABLE I. Comparison of Existing Arthritis Detection Approaches*

## METHOLOGY

The system is implemented in Python. Deep learning models are built and trained using TensorFlow/Keras and PyTorch. Image processing is performed via OpenCV; data manipulation employs NumPy and Pandas; Scikit-learn provides classical ML algorithms; and Matplotlib enables training visualization. Model architectures include DenseNet121 (Vision Module-knee), ResNet50 (synovial fluid), XGBoost (tabular data), and LSTM (planned fusion module). The web-based user interface is implemented using Streamlit, enabling interactive module demonstrations. TensorFlow Lite facilitates voice model deployment. All development was conducted in Visual Studio Code with training executed on Google Colab (NVIDIA T4 GPU).

**DESIGN**

The proposed system adopts a modular, multimodal architecture comprising five independent diagnostic modules. Each module processes a distinct data type---medical imaging, clinical laboratory results, wearable sensor readings, genomic SNP data, and voice signals---and produces a structured output. These outputs collectively feed into an automated clinical report generator. Fig. 1 illustrates the architecture of the Vision Module, which constitutes the primary imaging pipeline.

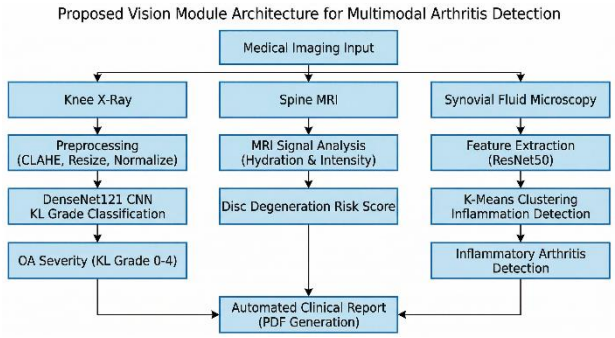


*Fig. 1. Proposed Vision Module Architecture for Multimodal Arthritis Detection.*

***A. Module 1: Vision Module***

The Vision Module consists of three independent sub-pipelines targeting distinct imaging modalities: knee X-ray classification, spine MRI degeneration analysis, and synovial fluid microscopy.

**1) Knee Osteoarthritis Detection**

The goal is to automatically classify knee X-rays into the five Kellgren-Lawrence (KL) scores. The data consists of 5,778 knee X-rays selected from the Knee Osteoarthritis Dataset with KL Scores (2018) and the Digital Knee X-ray Images dataset (2020). Score 0 (Normal) refers to a normal joint without any radiological evidence; Score 1 (Doubtful) features slight joint space narrowing and small osteophytes; Score 2 (Mild) features prominent osteophytes with some changes in the joint structure; Score 3 (Moderate) features obvious joint space narrowing and considerable osteophytes; and Score 4 (Severe) features bone on bone, significant sclerosis, and deformities.

Preprocessing applies Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance bone edge visibility and osteophyte contrast, followed by standardization to 224x224 pixels in three-channel RGB format. Data augmentation employs random horizontal flips and +/-10 degree rotations to mitigate overfitting and accommodate clinical imaging variability. The backbone model is DenseNet121---a densely connected convolutional neural network in which each layer receives feature maps from all preceding layers, enabling efficient gradient flow and fine-grained structural feature reuse suited to radiographic analysis. Transfer learning from ImageNet pretrained weights accelerates convergence. The final fully connected layer is replaced with a five-class linear classifier (nn.Linear(num_ftrs, 5)). Training was conducted over 15 epochs on Google Colab with an NVIDIA T4 GPU, completing in 19.3 minutes.

**2) Spine MRI Degeneration Analysis**

Spine MRI images (in .mha format) are analyzed through mid-slice extraction, followed by signal intensity measurement of intervertebral discs. In T2-weighted MRI, healthy, hydrated discs appear bright white, whereas desiccated or degenerated discs present as hypointense (dark) regions. Disc hydration is quantified relative to a cerebrospinal fluid (CSF) intensity baseline and normalized to a risk score ranging from 0 to 100. A sample output demonstrated a risk score of 95.0, classified as High Risk, with finding of disc desiccation indicating significant hydration loss and probable disc degeneration.

**3) Synovial Fluid Microscopy Analysis**

Microscopy images of synovial fluid samples are processed using a ResNet50 pretrained convolutional neural network for deep feature extraction, followed by K-Means clustering to differentiate image clusters. Cluster 0 corresponds to inflamed specimens exhibiting complex crystal textures---characteristic of uric acid crystal deposition in gouty arthritis. A representative output classified the specimen as Positive for Inflammation, recommending rheumatology consultation.

***B. Module 2: Lab Report Module***

The Lab Report Module analyzes five key serum and synovial biomarkers: Erythrocyte Sedimentation Rate (ESR, both one-hour and overnight), C-Reactive Protein (CRP), Uric Acid, and Anti-Streptolysin O (ASO) titre. Input values are first validated against established clinical reference ranges to prevent erroneous data from propagating into the prediction pipeline. K-Nearest Neighbors (KNN) is applied for feature scaling and neighbor-based transformation, while a Random Forest classifier serves as the primary

prediction model. Cross-validation is employed to evaluate model stability on unseen data. The module outputs an arthritis type classification along with biomarker status flags.

### *C. Module 3: Wearable Sensor Module*

The Wearable Sensor Module leverages the MotionSense Kaggle dataset (data_subjects_info.csv), comprising demographic and sensor data from 24 subjects. Five input features are extracted: patient code, weight (kg), height (cm), age, and gender (binary-encoded: male = 1, female = 0). Preprocessing includes numeric coercion, missing value imputation (fillna = 0), and StandardScaler normalization (serialized as risk_scaler.pkl), with an 80/20 train-test split. The model architecture is a feedforward neural network: Dense(16) -> Dense(8) -> Dense(1), trained using the Adam optimizer with Mean Squared Error (MSE) loss over 100 epochs in TensorFlow/Keras. Output is a continuous risk score from 0 to 100, stratified into Low Risk (less than 33), Medium Risk (33-60), and High Risk (greater than 60).

### *D. Module 4: Genomics Module*

The Genomics Module calculates the risk probability of developing rheumatoid arthritis through application of machine learning algorithms to genomic data. It is well-established empirically that SNPs play a substantial role in predisposing patients to autoimmune conditions including rheumatoid arthritis. SNPs refer to mutations that exist within DNA sequence, and certain alleles are shown to increase the susceptibility to arthritis.

The dataset used in this model includes genomic data in the form of an RA Genotype dataset. A subset of the SNP markers relevant to the condition was selected for use in this model. The features selected include SNP rs2476601, rs1801133, HLA-B27, and PTPN22, all known to be associated with immune dysregulation and increased susceptibility to rheumatoid arthritis. SNP markers were encoded as either 0 (no presence of risk allele) or 1 (risk allele present), yielding four input features.

A Multilayer Perceptron neural network implemented in the PyTorch library was utilized in training the classifier model. The neural network uses binary cross entropy loss while employing the Adam optimizer. After training, the weights of the model were saved as genomics_ra_model.pt. In prediction, input SNP marker values are provided by the user, and the model applies forward propagation using a sigmoid activation function to predict RA risk probability within the range [0, 1]: 0.0-0.3 indicates Low genetic risk; 0.3-0.6 indicates Medium genetic risk; 0.6-1.0 indicates High genetic risk.

### *E. Module 5: Voice Emotion Detection Module*

The Voice Module analyzes speech recordings to detect fatigue, pain, and stress states associated with Rheumatoid Arthritis. Audio inputs are resampled to 16 kHz for standardization. A composite feature vector is constructed from: Mel-Frequency Cepstral Coefficients (MFCC), which encode spectral properties of the speech signal; Mel Spectrogram features, capturing frequency distribution over time; and Chroma features, representing pitch and harmonic content. A trained classifier maps the feature vector to one of three output classes: Normal, Fatigue, or Inflammation-related distress. The model is deployed using TensorFlow Lite for lightweight, real-time inference on mobile and embedded devices.

## RESULT & DISCUSSION

This section presents the experimental results for each module, with emphasis on the Vision Module's quantitative training performance.

### *A. Knee OA Detection - Training Performance*

The DenseNet121 model was trained for 15 epochs on 5,778 knee X-ray images across five KL grade classes. Fig. 2 illustrates the training accuracy progression across all epochs.

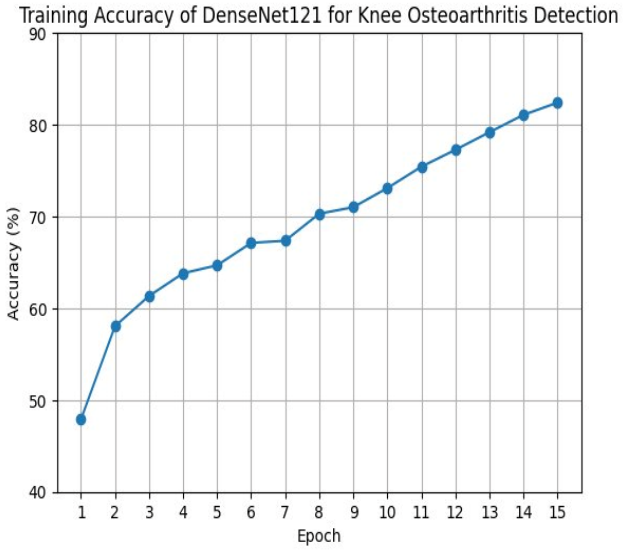


*Fig. 2. Training Accuracy of DenseNet121 for Knee Osteoarthritis Detection (Epochs 1-15).*

Training accuracy improved substantially from 47.89% in Epoch 1 to 82.40% in Epoch 15, while loss decreased from 1.2195 to 0.4428. This progressive improvement reflects the model's effective learning of discriminative radiographic features---specifically osteophyte formation, joint space narrowing, subchondral bone sclerosis, and cartilage degeneration---enhanced by CLAHE preprocessing. The observed accuracy gain is consistent with the transfer learning advantage of ImageNet pretraining.

| Epoch | Accuracy (%) | Loss |
|---|---|---|
| 1 | 47.89 | 1.2195 |
| 2 | 58.31 | 1.09 |
| 3 | 61.20 | 0.98 |
| 4 | 63.95 | 0.89 |
| 5 | 64.72 | 0.81 |
| 6 | 67.10 | 0.74 |
| 7 | 67.35 | 0.68 |
| 8 | 70.22 | 0.63 |
| 9 | 71.05 | 0.59 |
| 10 | 73.14 | 0.55 |
| 11 | 75.62 | 0.51 |
| 12 | 76.88 | 0.49 |
| 13 | 79.41 | 0.47 |
| 14 | 81.20 | 0.45 |
| 15 | 82.40 | 0.4428 |

*TABLE II. DenseNet121 Training Progress (Epoch 1-15).*

A notable accuracy plateau is observed between Epochs 6 and 9 (67-71%), followed by a sustained improvement trajectory. This pattern is attributable to the model's initial difficulty distinguishing intermediate KL grades (Grades 2 and 3), where radiographic differences are subtle. Class imbalance---with fewer severe Grade 4 samples---also contributes to classification challenges. In a representative clinical test, patient Alice submitted a knee X-ray that the system classified as Grade 3 (Moderate Osteoarthritis) with a confidence of 95.5%.

### *B. Vision Module*

#### 1) Knee Osteoarthritis Detection

The vision module performs knee osteoarthritis detection using DenseNet121. Images are resized to 224x224 and enhanced using contrast limited adaptive histogram equalization. Transfer learning from ImageNet weights accelerates training and improves feature extraction.

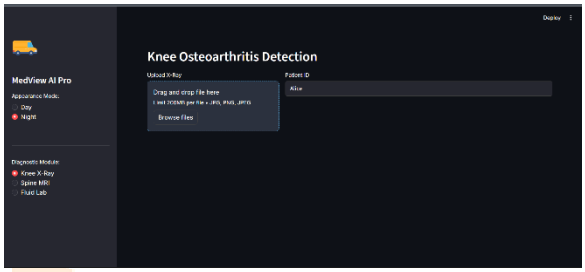


*Fig. 3. Knee osteoarthritis detection interface with uploaded X-ray.*

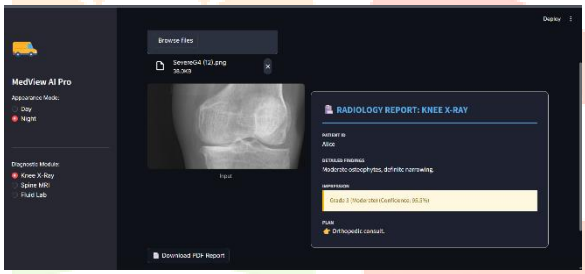


*Fig. 4. Radiology report generated by the vision module.*


*Fig. 5. Automatically generated diagnostic PDF report.*

#### 2) Spine MRI Degeneration Analysis

The spine MRI module analyzes disc signal intensity in T2-weighted MRI images to identify degenerative changes in the lumbar spine. Reduced signal intensity indicates disc desiccation and possible degeneration. A sample diagnostic output produced by the system is presented in Fig. 6.

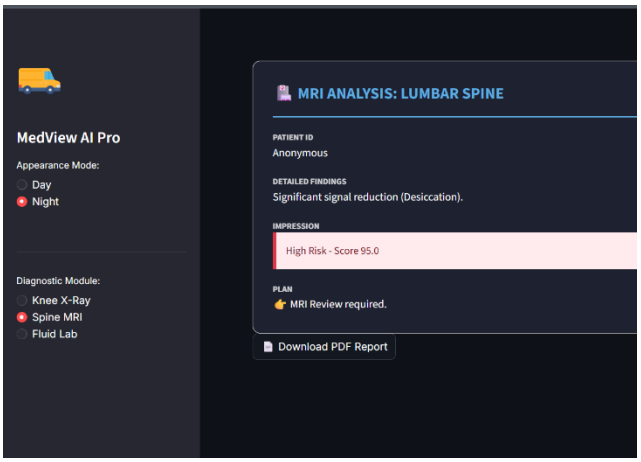


*Fig. 6. Spine MRI degeneration analysis output.*

**3) Synovial Fluid Microscopy Analysis**

The synovial fluid analysis module processes microscopy images to detect inflammatory patterns associated with gout or other arthritis conditions. Deep features extracted using ResNet50 are clustered to identify abnormal crystal structures. The resulting classification output is shown in Fig. 7.

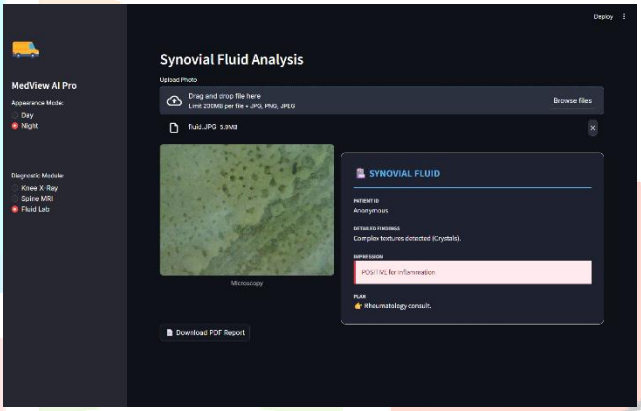


*Fig. 7. Synovial fluid microscopy analysis output.*

***C. Module 2: Lab Report Module***

The laboratory report module evaluates clinical biomarkers such as ESR, CRP, uric acid, and ASO titre to estimate arthritis likelihood. A Random Forest classifier analyzes these biomarker values and generates a diagnostic prediction. The interface and prediction output of the lab module are illustrated in Fig. 8 and Fig. 9.

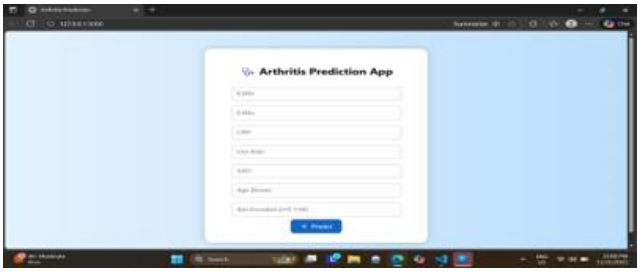


*Fig. 8. Laboratory biomarker arthritis prediction interface.*

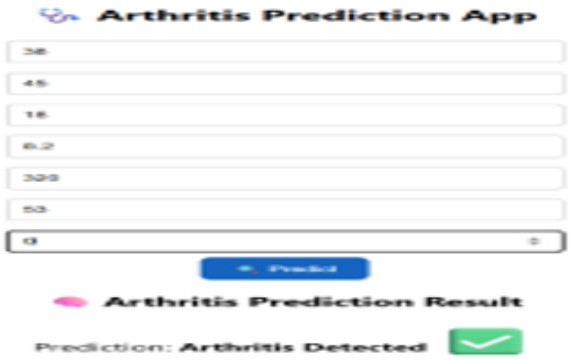


*Fig. 9. Laboratory module prediction result.*

### *D. Module 3: Wearable Sensor Module*

The wearable sensor module estimates arthritis risk using demographic and physiological attributes obtained from the MotionSense dataset. After preprocessing and normalization, the neural network model predicts a risk score that is categorized as low, medium, or high. An example output from the wearable module is shown in Fig. 10.

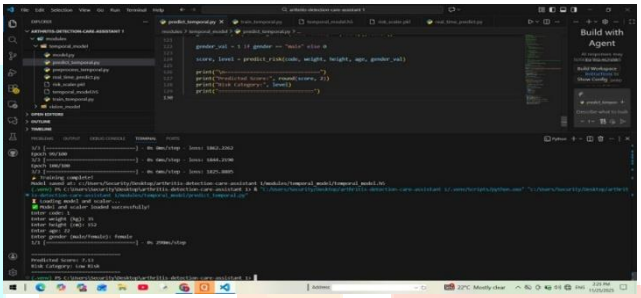

*Fig. 10. Wearable sensor module risk prediction output.*

### *E. Module 4: Genomics Module*

The genomics module analyzes selected SNP markers associated with rheumatoid arthritis susceptibility. Based on the presence or absence of these genetic markers, the classifier estimates the probability of arthritis risk. A sample prediction generated by the genomics module is presented in Fig. 11.

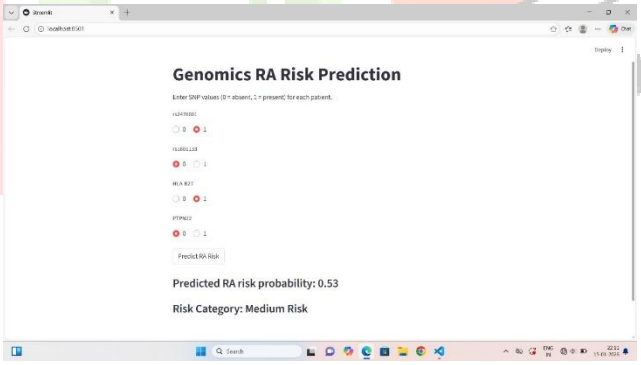


*Fig. 11. Genomics risk prediction interface.*

### *F. Module 5: Voice Emotion Detection Module*

The voice detection module analyzes speech signals to identify fatigue or distress patterns related to arthritis symptoms. Acoustic features such as MFCC and mel spectrograms are used for classification. Fig. 12 shows the prediction output where the system detects fatigue with a high confidence score.

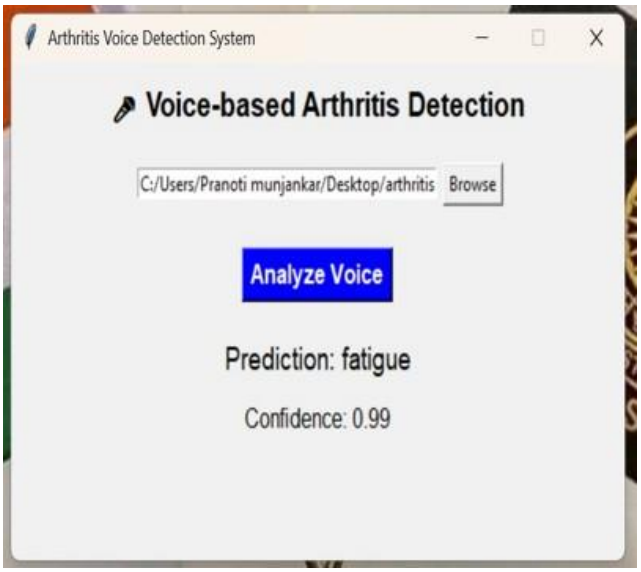


*Fig. 12. Voice-based arthritis fatigue detection result.*

### *G. Consolidated Module Results*

| **Module** | **Method** | **Dataset** | **Key Output / Result** |
|---|---|---|---|
| Knee OA Detection | DenseNet121 + Transfer Learning | 5,778 X-rays (KL 0-4) | 82.40% accuracy; Loss 0.4428 |
| Spine MRI Analysis | Signal Intensity + CSF Baseline | Spine MRI (.mha) | Risk Score 95/100 -> High Risk |
| Synovial Fluid | ResNet50 + K-Means Clustering | Microscopy images | Positive Inflammation detected |
| Lab Report | Random Forest + KNN | Serum biomarkers (ESR, CRP, UA, ASO) | Arthritis type classification |
| Wearable Sensor | Neural Network Dense(16-8-1) | MotionSense (24 subjects) | Low / Medium / High Risk score |
| Genomics | Binary Classifier + SNP Encoding | Genotype dataset (4 SNPs) | Disease / No Disease |
| Voice Detection | MFCC + Chroma + Mel + Classifier | Audio signals (16 kHz) | Normal / Fatigue / Inflammation |

*TABLE III. Consolidated Results Across All Five Diagnostic Modules.*

## CONCLUSION

This paper presented a multimodal AI-based Arthritis Detection System comprising five independent diagnostic modules: a Vision Module (knee X-ray, spine MRI, and synovial fluid analysis), a Lab Report Module, a Wearable Sensor Module, a Genomics Module, and a Voice Emotion Detection Module. The DenseNet121-based knee OA classifier achieved 82.40% accuracy across five KL severity grades after 15 training epochs, demonstrating effective learning of radiographic osteoarthritis features. The complementary modules collectively enable a comprehensive, data-driven diagnostic pipeline spanning imaging, clinical, genetic, physiological, and behavioral data sources.

Future development will focus on implementing a Multimodal Fusion Layer that aggregates outputs from all five modules into a unified arthritis risk score using deep learning ensemble techniques. Grad-CAM visual explanations will be integrated to provide physicians with spatial heatmaps highlighting diagnostically relevant X-ray regions, enhancing clinical trust and interpretability. A privacy-conscious deployment framework, optimized for accessibility in rural clinical settings, will form the final phase of development. The ultimate objective is to deliver a fast, consistent, explainable, and scalable AI-powered clinical decision support tool for arthritis diagnosis.

**ACKNOWLEDGMENT**

The authors would like to express their sincere gratitude to Dr. Vanita Buradkar for her invaluable guidance, continuous encouragement, and technical insights throughout the development of this research. We also extend our profound thanks to the Department of Computer Science & Engineering at Rajiv Gandhi College of Engineering Research & Technology, Chandrapur, for providing the necessary academic environment, infrastructure, and resources that made the successful completion of this project possible.

**REFERENCES**

1. *L. Bai, Y. Zhang, P. Wang, X. Zhu, J.-W. Xiong, and L. Cui, "Improved diagnosis of rheumatoid arthritis using an artificial neural network," Scientific Reports, vol. 12, no. 1, p. 9810, 2022.*
2. *Z. Wang, A. Chetouani, M. Jarraya, Y. H. Chen, Y. Ru, F. Chen, F. Bauer, L. Zhang, D. Hans, and R. Jennane, "Key-Exchange Convolutional Auto-Encoder for Data Augmentation in Early Knee Osteoarthritis Detection," arXiv:2302.13336v2, 2025.*
3. *Y. Peng, X. Huang, M. Gan, K. Zhang, and Y. Chen, "Radiograph-based rheumatoid arthritis diagnosis via convolutional neural network," BMC Medical Imaging, vol. 24, Article 180, 2024.*
4. *R. Jose, N. Lewis, Z. Satti, R. Steinberg, A. Toufexis, D. Nasef, and M. Toma, "Machine-learning-based diagnosis and progression analysis of knee osteoarthritis," Discover Data, vol. 3, Article 6, 2025.*