UNDERGRADUATE RESEARCH IN NATURAL AND CLINICAL SCIENCE AND TECHNOLOGY (URNCST) JOURNAL
Read more URNCST Journal articles and submit your own today at: https://www.urncst.com

PRIMARY RESEARCH OPEN ACCESS

Exploring the Feasibility of Applying Deep Learning for the Early Prediction of Arthritis

Jiaxuan Chen, HBSc Student [1]^, Xiangxuan Kong, HBSc Student [1]*^

[1] Faculty of Arts and Science, University of Toronto, Toronto, Ontario, Canada M5S 1A1

*Corresponding Author: noah.kong@mail.utoronto.ca
^All authors were equal contributors to this study.

Abstract

Introduction: Arthritis is one of the most common chronic diseases. Early detection of arthritis and its progression can facilitate early intervention measures, lowering disease severity in patients. As electronic health records (EHR) become more accessible, this study assesses whether general health information and arthritis-related questionnaires can be used in arthritis diagnosis, without the involvement of costly imaging methods. Therefore, we created deep learning (DL) and machine learning (ML) models to explore the feasibility of combining EHR and modern computational tools to diagnose arthritis.

Methods: A total of 782 arthritis patients and 4014 control patients were identified from the Osteoarthritis Initiative (OAI) – a ten-year-long observational study that included patient EHR in five time points. Six hundred variables were filtered by random forest classifier followed by manual filtering. Data were split properly to training, testing and validation set, and the training set was balanced. Sequential, nonsequential DL models, and five independent DL models for each time points were used. The accuracy, positive prevalence value (PPV), negative prevalence value (NPV), and area under curve (AUC), were assessed and compared with four classical ML models. SHAP (SHapley Additive exPlanations) summary analysis was also conducted.

Results: Sequential and non-sequential deep learning models showed accuracies of ~ 0.97, and the four classical machine learning approaches showed accuracies of above 0.9. High positive and negative predicted values (> 0.90) for all of the models suggested the potential clinical applicability of the model, while the SHAP analysis demonstrated its interpretability.

Discussion: We tested various models and showed the ability to use machine learning methods for early diagnosis of arthritis with EHR. The models can be used as a screening tool to select susceptible patients for confirmatory tests such as X-ray and MRI. Identification of early disease states could facilitate protective measures that slow disease progression.

Keywords: arthritis; deep learning; machine learning; artificial intelligence

Introduction

Arthritis is a condition caused by joint inflammation; it is the most prevalent chronic disease worldwide. There are over 100 types of arthritis, each capable of diminishing the quality of life of the affected population, with symptoms ranging from chronic joint pain to severe disability [1]. Arthritis is estimated to affect approximately 20% of Canadians [2], and the estimated risk for arthritis ranges as high as 47% in individuals older than 65 [3]. This prevalence is predicted to rise higher [4]. Currently, there are no satisfactory drugs that can stop arthritis progression or provide long-lasting symptomatic relief. The most effective therapy to improve quality of life is joint replacement, but they are typically expensive and harmful to perform in can be unsuccessful in elder patients [5]. One potential reason for the failure of interventions is that it may be too late to give the treatments to patients when the arthritis is already symptomatic. In fact, underdiagnosis of arthritis is significant. The pathological processes of different types of arthritis can start years before the onset of clinical symptoms, leaving a wide time window for early prediction and identification of risk factors [6]. However, asymptomatic and lightly symptomatic patients can go unnoticed until the late stages of progression, when major structural and functional alterations have occurred. Increasing evidence suggests that early recognition of arthritis is critical. Firstly, lifestyle adaptation and preventative treatments could slow symptom progression. Furthermore, researchers could better characterize biomarkers associated with early stages of the disease [7].

Many databases have been established to investigate the onset and progression of arthritis, including The National Data Bank for Rheumatic Diseases, the Osteoarthritis Initiative (OAI), and American Rheumatism Association Medical Information System [8–10]. They provide essential

platforms for investigating different biomarkers and the pathophysiology of different types of arthritis. As computational methods evolve, research has suggested a potentially promising role for AI systems in the diagnosis of arthritis. Several studies have used machine learning models for risk prediction and pain assessment for osteoarthritis and rheumatoid arthritis, involving both statistical and machine learning (ML) approaches [9, 11].

With the continuous improvement of data collection, large collections of longitudinal electronic health records (EHR) are becoming available for clinical research. Numerous Many risk factors of arthritis identified by previous studies (e.g., obesity, joint injuries, age) can be routinely captured in patients' EHRs [12]. By using EHR as a diagnostic tool, it can facilitate the early detection of arthritis onset, without the involvement of complicated, and costly tests. Many ML methods are commonly applied to clinical data, such as random forests, K-nearest neighbour, and support vector machine [11]. However, these methods are often not suited to EHR data due to the high number of variables and the presence of missing data [13]. Deep learning (DL) is a subtype of ML that has gained popularity in different health domains due to its ability to process complex data. With underlying algorithms that mimic human thinking processes, it enables computers to learn and evolve in an astonishing fashion [14]. Therefore, this study comprehensively explored the feasibility of DL models, compared to ML models, by using large collections of real-world EHR data from The Osteoarthritis Initiative databases - a multicenter, ten-year observational study of men and women, sponsored by the National Institutes of Health. Successes in the models could support early diagnosis of arthritis with EHR. The potential usage of AI in clinical settings involves acting as a screening tool to select susceptible patients for confirmatory tests. Identification of early disease states could facilitate protective measures that reduce disease morbidity.

Methods

Data Description

The OAI longitudinal dataset consists of 5 time points collected over a span of 8 years, involving 4796 participants [15].

Data Processing

Feature Selection

In the initial step of feature selection, we used random forest classifier to extract the important variables, and then manually eliminated irrelevant variables from the dataset [16, 17]. The variables were scrutinized based on their domain relevance to arthritis as well as previous literature indicating their significance or lack thereof. This process reduced the dimensionality of the dataset and helped in focusing on the most pertinent features for disease prediction.

Our dataset, initially in a 3D array format (patients, time steps, features), was reshaped into a 2D format, aligning each patient-time step combination with a corresponding feature. The patient's ID column was dropped before training to avoid bias.

Data Imputation & Normalization

Features were classified into categorical and numerical types. Numerical features underwent median imputation and standardization, while categorical features were imputed with the most frequent category. This approach was facilitated by Scikit-learn's Pipeline and ColumnTransformer functionalities.

Data Sampling

The data was randomly split for training (80%) and testing (20%). In some experiments, 10% or 20% of the training data is split out for validation during training. The outcome of OA disease diagnosis (Y/N) served as the dependent variable and was used as ground truth.

Oversampling was applied using SMOTE (Synthetic Minority Over-sampling Technique), to address class imbalance by augmenting the minority class [18]. It operates by creating synthetic samples rather than replicating existing ones. For a given minority class sample (in this case, the diseased participants), SMOTE selects one or several of its nearest neighbors from the same class, then synthesizes a new sample that is a linear interpolation between the chosen sample and its neighbors. This process involves randomly choosing a point along the line segment connecting the sample under consideration with its selected neighbor(s).

Model Training

Robust Neural Network Models: Non-sequential Analysis

Data across 2009-2016 was reshuffled to produce non-sequential data. The model's performance is contingent on several hyperparameters, which were carefully chosen.

1) Number of Neurons: The hidden layer comprised 64 neurons, providing the model with sufficient complexity to capture patterns in the data without being overly prone to overfitting.
2) Activation Functions: The ReLU activation function was used in the hidden layer.
3) Dropout Rate: Set at 0.5.
4) Loss Function and Optimizer: The model employed binary cross-entropy as the loss function, apt for binary classification tasks, and used the Adam optimizer.
5) Training Parameters: The model was trained over 50 epochs with a batch size of 32. These parameters were selected to ensure sufficient training for convergence while maintaining computational efficiency.
6) A validation split of 20% was used during training, allowing for the monitoring of model performance on unseen data and aiding in the prevention of overfitting.
7) Undersampling was applied to the training data.

Longitudinal Analysis Using Traditional Neural Networks

Neural network model was trained by data from the first 4 time points and was tested by data at the end of 2016. The model used was a simple feedforward neural network with one hidden layer of 64 neurons, a dropout layer for regularization, and an output layer with sigmoid activation. The model was compiled with binary cross-entropy loss and trained over 50 epochs with a batch size of 32. Details of model parameters are the same as the robust neural network model (method 2.4.1). Undersampling was applied to the training data set.

Individual Deep Learning Model for Data at Single Time Points

Individual deep learning models were developed for data at each specific time point, focusing on a detailed analysis of the data's temporal characteristics. The model applied a batch size of 32 and 50 epochs.

Traditional Machine Learning Methods

We used other mainstream machine learning models including Random Forest [19], XGBoost (XGB) [20], Support Vector Machine (SVM) [21], and K-Nearest Neighbors [22] to fit the dataset collected in 2009 and tested their performance on testing data. The train-test split strategy aligns with the practices above (80% vs 20%)

Model Evaluation Methods

To rigorously assess the performance of the trained models, several evaluation metrics were employed.

Accuracy (PPV, NPV, Accuracy, Sensitivity and Specificity)

Accuracy was the fraction of correct predictions made by the model. It was measured by dividing the number of correct predictions by the total number of predictions. It serves as a straightforward evaluation metric, especially when the class distribution is balanced. PPV, NPV, sensitivity and specificity of models were also assessed.

Confusion Matrices

In our study, the confusion matrix was utilized as a critical tool for evaluating the performance of a classification model in differentiating between arthritis and non-arthritis populations. This matrix presents a detailed breakdown of the model's predictions, categorized into four key segments: True Positives (TP), where the model correctly identifies individuals with arthritis; True Negatives (TN), where it correctly recognizes non-arthritis individuals; False Positives (FP), cases where non-arthritis individuals are mistakenly classified as having arthritis; and False Negatives (FN), where individuals with arthritis are incorrectly classified as non-arthritic.

ROC and AUC

The Receiver Operating Characteristic (ROC) curve plots the true positive rate against the false positive rate at various decision thresholds [23]. The Area Under the Curve (AUC) summarizes the ROC curve into a single value, indicating the model's ability to distinguish between the classes. A higher AUC indicates better model performance.

SHAP (SHapley Additive exPlanations)

At the end, we attempted to apply SHAP, is a significant advancement in machine learning interpretability, leveraging Shapley values from cooperative game theory to elucidate our model outputs [24]. It systematically quantified the contribution of each feature to a model's prediction, enabling precise interpretation even in complex models such as deep neural networks or ensemble methods. This granular, feature-level insight afforded by SHAP is indispensable for validating the internal mechanics of models, ensuring transparent, fair decision-making. We could use it to explore the importance of each variable in our data analysis model.

Results

Traditional Neural Networks

Non-Sequential Analysis

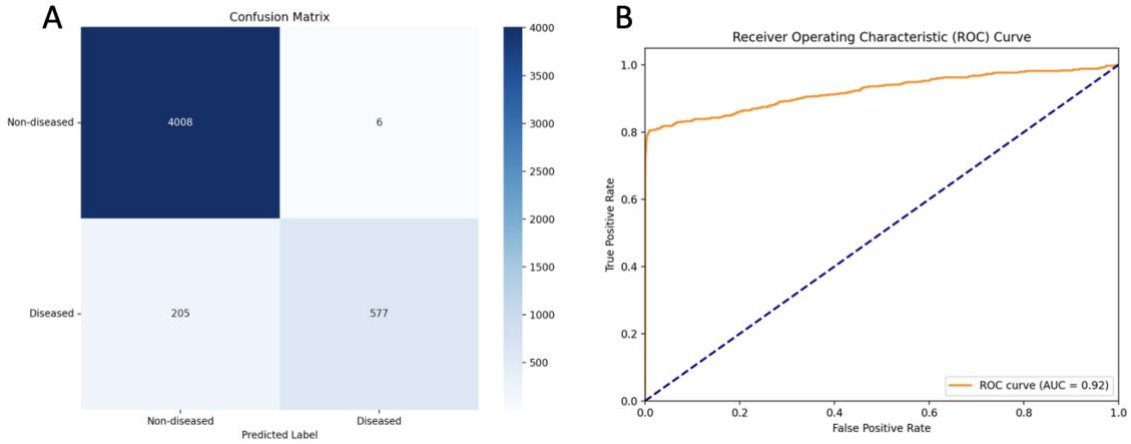

Figure 1. Performance of Traditional Neural Network Model Employing Non-Sequential Analysis. A. Confusion matrix visualizes the performance of the classification model. The matrix compares the actual vs. predicted labels, showcasing the number of true positives (top left), true negatives (bottom right), false positives (top right), and false negatives (bottom left). B. ROC graph represents the diagnostic ability of the binary classifier. The curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. All subfigures are made with Python (https://www.python.org/).

The neural network model in Figure 1 had a positive predictive value (PPV) of 0.99, a negative predictive value (NPV) of 0.95, a sensitivity of 0.73, and a specificity of 0.95. Area under curve was 0.92.

Sequential Analysis

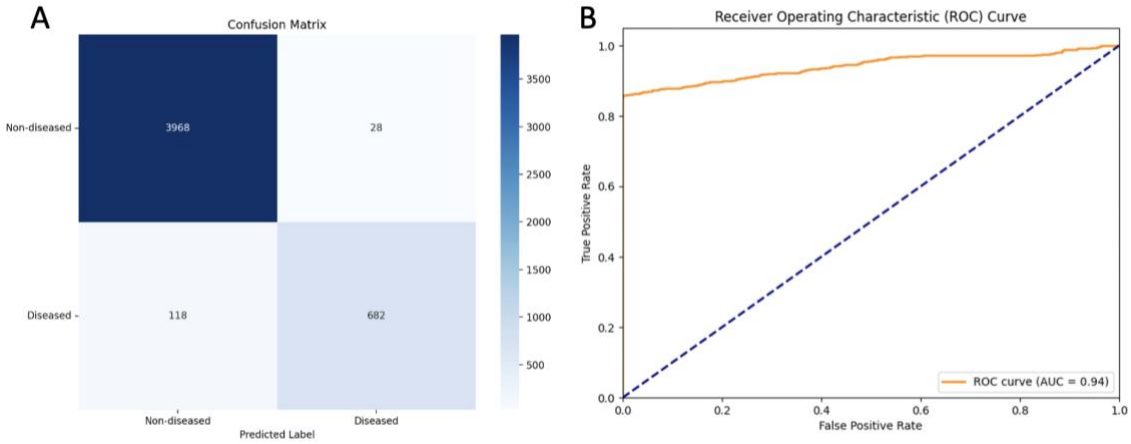

Figure 2. Performance of Traditional Neural Network Model Employing Sequential Analysis. A. Confusion matrix visualizes the performance of the classification model. The matrix compares the actual vs. predicted labels, showcasing the number of true positives (top left), true negatives (bottom right), false positives (top right), and false negatives (bottom left). B. ROC graph represents the diagnostic ability of the binary classifier. The curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. All subfigures are made with Python (https://www.python.org/).

The neural network model in Figure 2 has a positive predictive value (PPV) of 0.96, a negative predictive value (NPV) of 0.97, a sensitivity of 0.85, and a specificity of 0.97.

Individual Time Points Analysis

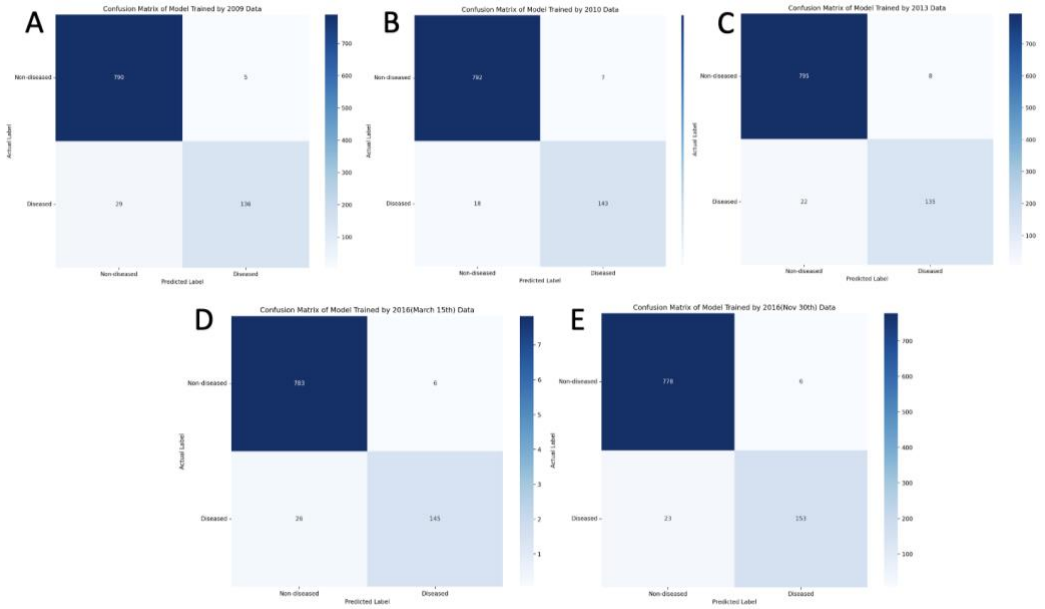

Figure 3. Confusion Matrices for Individual Neural Network Models. This graph demonstrates the effect of learning through epochs. A. Confusion matrix of the neural network model performance on the data of 2009. B. Confusion matrix of the neural network model performance on the 2010 data. C. Confusion matrix of the neural network model performance on the 2013 data. D. Confusion matrix of the neural network model performance on the 2016 (March 15th) data. E. Confusion matrix of the neural network model performance on the 2016 (November 30th) data. All subfigures are made with Python (https://www.python.org/).

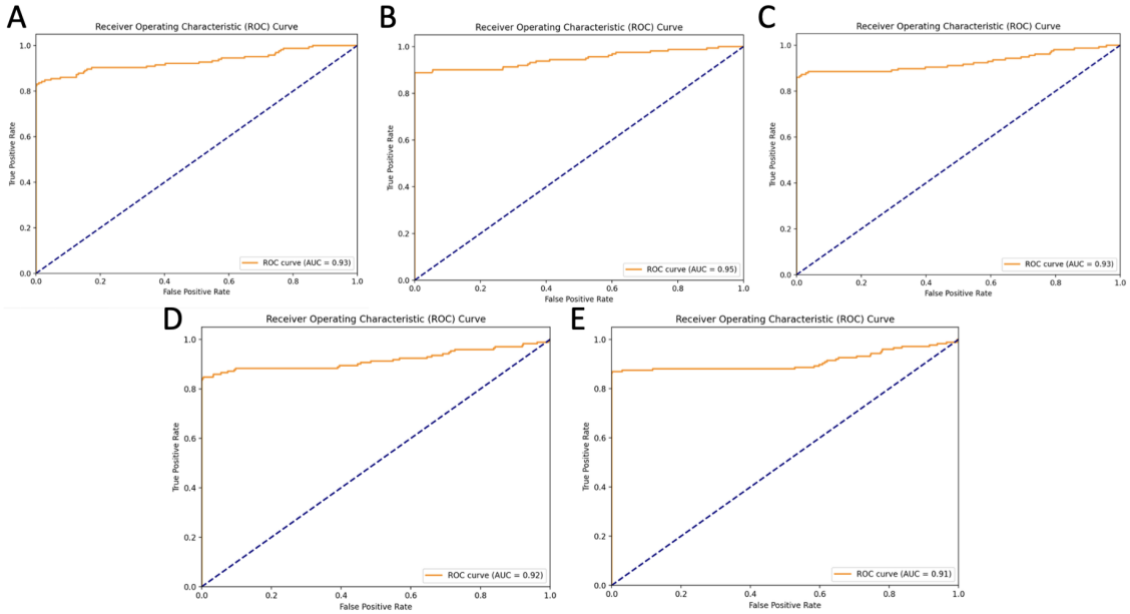

Figure 4. Receiver Operating Characteristic (ROC) Curves for Individual Neural Network Models. This figure presents a series of five ROC curves labeled from A to E (5 time points). Each panel displays the trade-off between the true positive rate (TPR) and false positive rate (FPR) for a binary classifier at various threshold settings. The ROC curves are plotted with the TPR on the y-axis against the FPR on the x-axis. The diagonal dashed line represents the line of no-discrimination, indicating a classifier with no better accuracy than random chance. The area under each ROC curve (AUC) is provided. All subfigures are made with Python (https://www.python.org/).

Table 1. Performance of Individual Deep Learning Model Trained on Data at Each Time Point

Time Point	Accuracy (PPV)	Accuracy (NPV)	Accuracy	Sensitivity	Specificity
May 8th 2009	0.96	0.97	0.97	0.83	0.96
February 26th 2010	0.98	0.98	0.98	0.89	0.98
December 11th 2013	0.94	0.97	0.96	0.86	0.97
March 15th 2016	0.93	0.97	0.95	0.85	0.97
November 30th 2016	0.96	0.97	0.97	0.86	0.97

All 5 models had accuracies of above 0.95 (Figure 3; Table 1). Area under curve is over 0.9 for all 5 individual neural network models (Figure 4). The models developed in this study exhibited a consistent pattern of lower sensitivity of around 80% coupled with higher specificity, this is especially observed in the individual time analysis (Table 1).

Robust Analysis Using Common Machine Learning Model

Table 2. Performance of Mainstream Machine Learning Model

Model	Accuracy	Sensitivity	Specificity
Random Forest	0.97	0.82	0.96
XGBoost	0.97	0.82	0.96
K-Nearest Neighbors (k = 300)	0.92	0.75	0.95
Support Vector Machine	0.95	0.73	0.95

Notes: Here, we used the data for 2009. Accuracy indicates the percentage of correct predictions made by the model. All the models are built in Python (https://www.python.org/).

As described in Table 2, accuracies of traditional machine learning models were all above 0.91 (0.92 to 0.97), with a sensitivity ranking ranging from 0.73 to 0.82, and specificity ranging from 0.95 to 0.96.

SHAP Analysis

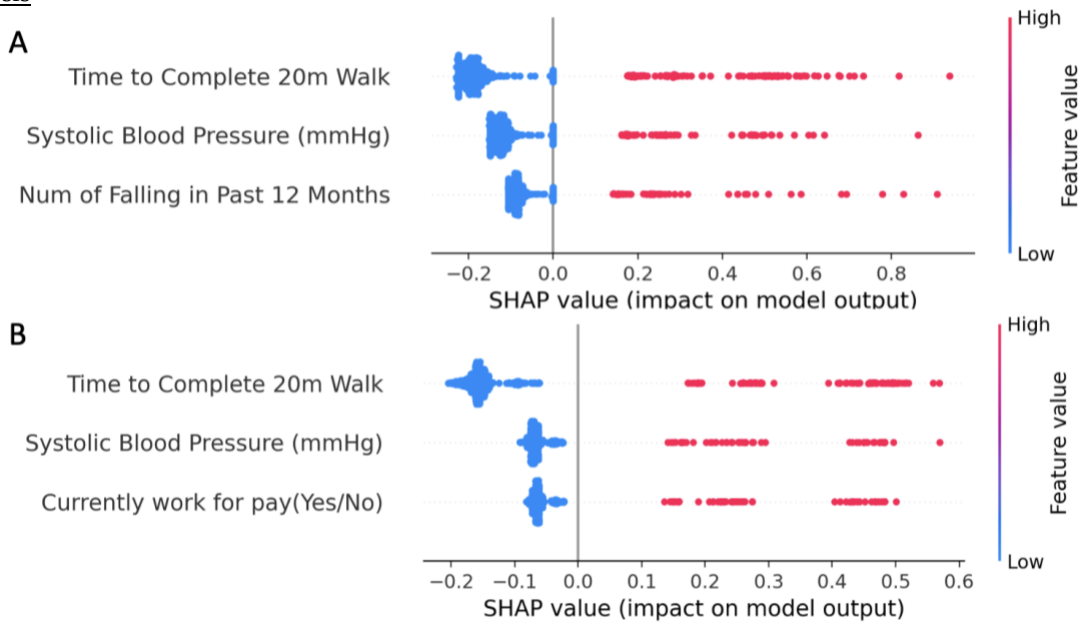

Figure 5. SHAP value graphs for the Top 3 Variables of DL and ML Models. A. SHAP value graph for the top 3 important variables in the neural network model trained by the data of November 2016. B. SHAP value graph for the top 3 important variables in the random forest model trained by the data of 2009. All subfigures are made with Python (https://www.python.org/).

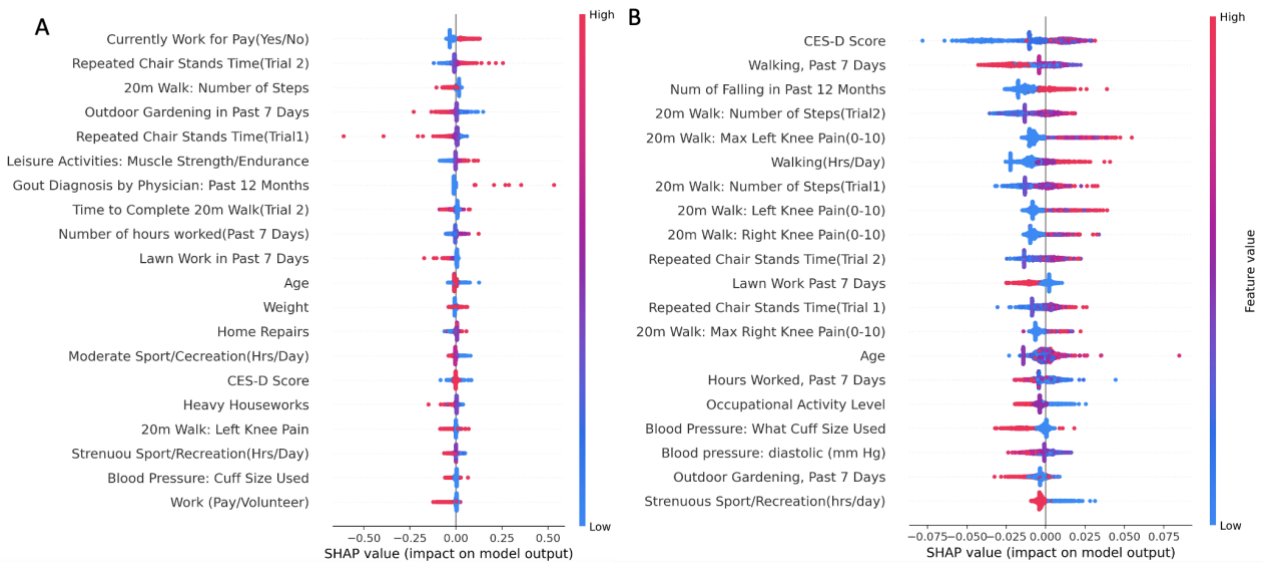

Figure 6. SHAP value graph for other important variables in DL and ML models. Eleven variables exist in both graphs. A. SHAP value graph for the other important variables in the neural network model trained by the data of November 2016. B. SHAP value graph for the other important variables in the random forest model trained by the data of 2009. All subfigures are made with Python (https://www.python.org/).

The color of the points indicates the SHAP value of the feature, with blue representing lower or less and pink representing higher or more. For example, a higher value of "Time to complete 20m walk" (pink points) tends to increase the model's chance of making arthritis prediction, while a lower value (blue points) tends to decrease it. Metrics from 20m walk examinations, systolic blood pressure and intensities of activities were shown to have particular importance in model decisions (Figure 5, Figure 6).

Discussion

In this study, we evaluated the feasibility of using deep learning models to identify arthritis patients from the OAI database and compare their effectiveness with classic ML models such as Random Forest, XGBoost, K-Nearest Neighbors, and Support Vector Machine. The neural network model employed in this study represented deep learning models widely recognized for their efficacy in pattern recognition and predictive analysis [19]. Neural networks consist of layers of interconnected nodes or neurons, where each connection represents a weight that is adjusted during the learning process [25]. In binary classification tasks, such as predicting the onset of arthritis from clinical data, neural networks learn to map input features to a binary outcome through a series of non-linear transformations.

During the training of neural networks, we approached the data in non-sequential (Figure 1), sequential ways (Figure 2), and by separating models for each individual timepoints (Figure 3). The non-sequential analysis was performed after randomly assorting data from different years, this is an approach that discards longitudinal patterns, allowing the model to predict only based on individual time points. On the other hand, sequential analysis used the first four time points for training and the last for testing. The model treated data from different time points as independent observations, aiming to capture a generalized pattern of arthritis onset from various stages of patient data. The study provided a logical separation, ensuring that the model was tested on unseen data from a different time point, which could be indicative of its generalizability. Lastly, we also generated neural network models by using data from each individual time point. Even though the sample size is much smaller than the other two methods, this prevents repetitive information from the same patient. The sequential analysis we conducted suggests that creating a model capable of predicting the onset of osteoarthritis years in advance is feasible, as our model demonstrated comparable performance to other robust neural network models (Figure 1, 2, 3). However, this study does not encompass longitudinal analysis, where the model would predict the onset of arthritis in the same patient based on their historical data, for two primary reasons: 1) The diagnosis results at each time point were self-reported, leading to potential inaccuracies in the longitudinal information associated with these results. 2) There was a lack of consistency in the self-reported diagnosis results across the five examinations, posing challenges for the model's accurate classification. For example, some participant had arthritis during several years in the middle of the longitudinal study, but not at the beginning or the last several years.

Random Forest, XGBoost (XGB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were exploited. Random Forest is an ensemble learning method known for its robustness and effectiveness in handling both

regression and classification tasks [16, 17]. XGBoost, standing for Extreme Gradient Boosting, is an optimized gradient boosting library that excels in speed and performance [21]. SVM is a versatile algorithm capable of performing linear and non-linear classification, outlier detection, renowned for its effectiveness in high-dimensional spaces [20]. KNN, simple yet effective, is a non-parametric algorithm that classifies data based on the majority class of its nearest neighbors, making it intuitive and useful for classification and regression problems in various settings [22].

To our surprise, ML models and deep learning models preformed equally well. The accuracies of over 95% indicated the huge potential for machine learning model models to diagnose arthritis using electronic health data. Suggestions of arthritis risks in real-world clinical settings without using imaging results can potentially save costs for patients who did not have the disease, offer potential risk indicators for the development of arthritis, and redistribute resources for health care providers. Given the clinical setting, XGBoost or SVM would generally be the better choices in scenarios involving high-dimensional data. XGBoost is particularly advantageous for its performance with large, complex datasets, while SVM is suitable for cases where the dataset is not excessively large and the class separation is clear. Since the diagnosis is fundamentally a binary classification problem, both algorithms will work well. The final decision would depend on the specific dataset characteristics and clinical requirements. In some cases where the interpretability is more important than the prediction accuracy, the random forest model may have an advantage.

Besides, the ROC curves with AUC values between 0.9 and 0.95 (Figure 4) showed that the individual model had an around 90% chance to detect discriminate between a randomly chosen diseased participant and a randomly chosen healthy participant (Figure 1B, 2B & 4). Furthermore, the PPV was lower than the NPV in almost all models, indicating a potential bias during the training phase, even though all the training data was well balanced using oversampling or under-sampling (Figure 1A & 2A) (Table 1 & 2). A lower Positive Predictive Value (PPV) compared to Negative Predictive Value (NPV) signifies a greater accuracy in identifying negative instances over positive ones. This discrepancy often arises due to a higher incidence of false positives, where the test or model incorrectly labels negative cases as positive. This can be particularly pronounced in situations where the condition's prevalence is low, leading to a situation where, despite high sensitivity and specificity, the PPV remains low due to the scarcity of true positive cases.

The models developed in this study exhibited a consistent pattern of lower sensitivity of around 80% coupled with higher specificity, this is especially observed in the individual time analysis (Table 1). This trend suggests that while the models are highly effective in correctly identifying healthy patients, they are somewhat less adept at detecting diseased patients. Such a disparity could be attributed to various factors inherent in the dataset and the model architecture. It is possible that the dataset contains more distinct and consistent features for non-arthritic cases, leading to higher specificity. In contrast, the heterogeneity in manifestations of arthritis might not be fully captured, affecting sensitivity. Also, imbalance in the dataset, with a possible under-representation of arthritic cases, could have biased the model towards the majority class, further contributing to this discrepancy.

Notably, the K-nearest Neighbors model shows a relatively low accuracy and sensitivity score compared with other machine learning model (Table 2). This may be due to overfitting caused by the oversampling of the dataset in the high-dimensional space. KNN models, as mentioned in method section, plot the test data with the training data, and study the neighboring data points to make classifications, which resembles how SMOTE oversamples the data. Therefore, the oversampling strategy performed by SMOTE might overwhelm the KNN model.

The SHAP value graphs from both the random forest and our specialized deep learning model (Figures 5 & 6) demonstrated the interpretability of our models by revealing similar influential variables that steer the predictions, despite the models being trained on distinct datasets and utilizing different learning algorithms. These graphs serve not only to validate the models' interpretability but also to confirm the consistency of variable importance across diverse analytical methods. For instance, systolic blood pressure, known to have a positive correlation with osteoarthritis onset, maintains a positive impact on the disease onset predictions made by both models (Figure 5A & B) [26]. In contrast, diastolic blood pressure is inversely related to the model's predictions (Figure 6B). There were other significant variables, such as those related to the 20-meter walk performance, the frequency of muscle-strengthening leisure activities, and the intensity of household chores, whose relevance to the models' decisions is both consistent and logical (Figure 6).

Our study does have several limitations. First, our results were not validated with an external EHR data set. Although OAI is a large, real-world clinical database, the study population may not be representative of the general arthritis population. In fact, a higher proportion of participants in this study have a risk for osteoarthritis than other types of arthritis, which can make the model predicting powers biased to osteoarthritis. Second, this study is a secondary analysis of observational EHR data, we relied on the data quality of the database. Missing data are present and documentation errors could occur, many variables are the subjective responses from patients and bias could occur through the timepoints. Data imputation process might neglect that and therefore imputing noisy data that influence the model performance. Thirdly, we relied on patient reports on whether they are diagnosed with arthritis. Some of the data points that are treated as non-diseased could have been

diseased but not diagnosed. Fourthly, the PPV was overall lower than the NPV (Table 1). The reason for the problem should be the scarcity of the data for diseased participants, leading to bias in the oversampling algorithm. Other studies in previous studies reduce this confounding variable by implementing assessing methods based on imaging results to robustly categorize patients' disease states. Lastly, the training and testing data include results from the same patients at different time points. Despite that we dropped the column for patient's ID, sequential and non-sequential deep learning models might "identify" patient through feature characteristics. This might result in biased accuracy for the model. Nevertheless, EHR-based study shows promise and potential in real world settings.

Conclusions

Arthritis is often underdiagnosed; many people do not take caring actions before symptoms get unmanageable. This study systematically explored the feasibility of using different AI algorithms including DL and ML models. Our study identified no significant difference in accuracy among different predicting models, but we compared the important features that each model used, which could be used to help reveal future research directions. The current models present the potential to identify patients with arthritis, it can be used as a screening technique to select susceptible patients for confirmatory studies. Future work shall focus on increasing model accuracy by increasing data size for both training and testing groups and include some EHR from the general public to reduce model bias. Other data sets should be used to validate the prediction models and a more knowledge-driven approach for variable filtering should be applied to increase the interpretability of the model.

List of Abbreviations Used

AI: artificial intelligence
AUC: area under curve
CES-D Score: center for epidemiologic studies depression scale score
DL: deep learning
EHR: electronic health record
FN: false negative
FP: false positive
KNN: k-nearest neighbors
ML: machine learning
MRI: magnetic resonance imaging
NPV: negative predictive value
OAI: osteoarthritis initiative
PPV: positive predictive value
ROC: receiver operating characteristic
SHAP: shapley additive explanations
SMOTE: synthetic minority over-sampling technique
SVM: support vector machine
TN: true negative
TP: true positive
XGB: XGBoost

Conflicts of Interest

The authors declare that they have no conflict of interests in publication.

Ethics Approval and/or Participant Consent

The use of the OAI data is approved by the National Institute of Mental Health Data Archive.

Authors' Contributions

JC: Contributed to study design, assisted with the collection of data, conducted manual variable filtering, and drafted the abstract, introduction, and discussion section of the manuscript.
XK: Made contributions to the design of the study, collected data, created the machine learning models, and drafted the method, result, and discussion section of the manuscript.

Acknowledgements

Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive (NDA). OAI is a collaborative informatics system created by the National Institute of Mental Health and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to provide a worldwide resource to quicken the pace of biomarker identification, scientific investigation and OA drug development. Dataset identifier(s): [NIMH Data Archive Collection ID(s) or NIMH Data Archive Digital Object Identifier (DOI)]. This study was also supported by Randa Mudathir, a mentor from the Undergraduate Research in Natural and Clinical Science and Technology (URNCST) Journal, we thank her for her advice and feedback during our planning and writing of this manuscript.

Funding

This study was not funded.

References

[1] Senthelal S, Li J, Ardeshirzadeh S, Thomas MA. Arthritis. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2023 [cited 2023 Nov 26]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK518992/
[2] Chapter 1: Life with arthritis in Canada: a personal and public health challenge – What is arthritis and how common is it? [Internet]. 2010 [cited 2023 Nov 26]. Available from: https://www.canada.ca/en/public-health/services/chronic-diseases/arthritis/life-arthritis-canada-a-personal-public-health-challenge/chapter-one-what-is-arthritis-and-how-common-is-it.html
[3] The truth about arthritis - Prevalence & impact [Internet]. Arthritis Society Canada. [cited 2023 Nov 26]. Available from: https://arthritis.ca/about-arthritis/what-is-arthritis/the-truth-about-arthritis

[4] Arthritis related statistics [Internet]. Centers for Disease Control and Prevention. 2023 [cited 2023 Nov 26]. Available from: https://www.cdc.gov/arthritis/data_statistics/arthritis-related-stats.htm
[5] Weinstein AM, Rome BN, Reichmann WM, Collins JE, Burbine SA, Thornhill TS, et al. Estimating the burden of total knee replacement in the United States. J Bone Joint Surg Am. 2013 Mar 6;95(5):385–92. https://doi.org/10.2106/jbjs.l.00206
[6] Turkiewicz A, Gerhardsson de Verdier M, Engström G, Nilsson PM, Mellström C, Lohmander LS, et al. Prevalence of knee pain and knee OA in southern Sweden and the proportion that seeks medical care. Rheumatology. 2015 May 1;54(5):827–35. https://doi.org/10.1093/rheumatology/keu409
[7] Bortoluzzi A, Furini F, Scirè CA. Osteoarthritis and its management - Epidemiology, nutritional aspects and environmental factors. Autoimmunity Reviews. 2018 Nov 1;17(11):1097–104. https://doi.org/10.1016/j.autrev.2018.06.002
[8] Michaud K. The national data bank for rheumatic diseases (NDB). Clin Exp Rheumatol. 2016;34(5 Suppl 101):S100–1. https://pubmed.ncbi.nlm.nih.gov/27762196/
[9] Joseph GB, McCulloch CE, Nevitt MC, Neumann J, Gersing AS, Kretzschmar M, et al. Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: Data from the osteoarthritis initiative. J Magn Reson Imaging. 2018 Jun;47(6):1517–26. https://doi.org/10.1002/jmri.25892
[10] Fries JF, Hess EV, Klinenberg J. A standard database for rheumatic diseases. Arthritis Rheum. 1974; 17(3):327–36. https://doi.org/10.1002/art.1780170319
[11] Momtazmanesh S, Nowroozi A, Rezaei N. Artificial intelligence in rheumatoid arthritis: Current status and future perspectives: A state-of-the-art review. Rheumatol Ther. 2022 Oct;9(5):1249–304. https://doi.org/10.1007/s40744-022-00475-4
[12] Arthritis risk factors [Internet]. Centers for Disease Control and Prevention. 2023 [cited 2023 Nov 26]. Available from: https://www.cdc.gov/arthritis/basics/risk-factors.htm
[13] Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. JAMA Network Open. 2019 Mar;2(3):e190606. https://doi.org/10.1001/jamanetworkopen.2019.0606
[14] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE Journal of Biomedical and Health Informatics. 2018 Sep;22(5):1589–604. https://doi.org/10.1109/jbhi.2017.2767063
[15] The Osteoarthritis Initiative [Internet]. National Institutes of Health. 2012 [cited 2024 Feb 29]. Available from: https://nda.nih.gov/oai
[16] Breiman L. Random forests. Machine learning. 2001;45(1):5-32. https://doi.org/10.1023/A:1010933404324
[17] Rigatti SJ. Random forest. Journal of Insurance Medicine. 2017 Jan 1;47(1):31-9. https://doi.org/10.17849/insm-47-01-31-39.1
[18] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. https://doi.org/10.1613/jair.953
[19] Sarker IH. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science. 2021 Aug 18;2(5):420. https://doi.org/10.1007/s42979-021-00815-1
[20] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2016. https://doi.org/10.1145/2939672.2939785
[21] Noble WS. What is a support vector machine? Nature Biotechnology. 2006 Dec 1;24(12):1565-7. https://doi.org/10.1038/nbt1206-1565
[22] Peterson LE. K-nearest neighbor. Scholarpedia. 2009 Feb 21;4(2):1883. http://doi.org/10.4249/scholarpedia.1883
[23] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997 Jul 1;30(7):1145-59. https://doi.org/10.1016/S0031-3203(96)00142-2
[24] Van den Broeck G, Lykov A, Schleich M, Suciu D. On the tractability of SHAP explanations. Journal of Artificial Intelligence Research. 2022 Jun 23;74:851-86. https://doi.org/10.1613/jair.1.13283
[25] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436-44. https://doi.org/10.1038/nature14539
[26] Zhang YM, Wang J, Liu XG. Association between hypertension and risk of knee osteoarthritis: A meta-analysis of observational studies. Medicine. 2017 Aug;96(32). https://doi.org/10.1097%2FMD.0000000000007584

Article Information
Managing Editor: Jeremy Y. Ng
Peer Reviewers: Randa Mudathir, Karan Malhotra
Article Dates: Received Dec 03 23; Accepted Feb 20 24; Published Mar 18 24

Citation
Please cite this article as follows:
Chen J, Kong X. Exploring the feasibility of applying deep learning for the early prediction of arthritis inhibitors and angiotensin receptor blockers (ARBs). URNCST Journal. 2024 Mar 18: 8(3).
https://urncst.com/index.php/urncst/article/view/562
DOI Link: https://doi.org/10.26685/urncst.562

URNCST Journal
"Research in Earnest"

Funded by the Government of Canada | Canada