Classification of Osteoarthritis Disease Severity Using Adaboost Support Vector Machines

T R Adyalam¹, Z Rustam¹ and J Pandelaki²

¹Department of Mathematics, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia
²Department of Radiology, Universitas Indonesia, Kampus UI Salemba, Jakarta 10430, Indonesia

Abstract. Osteoarthritis (OA) is a condition when the joint is painful due to mild inflammation that arises due to friction of the ends of the joint bone. OA is the most chronic disease and joint disability in elderly people. One way to prevent this disease is to do early detection using machine learning for classification. In this study, it was used Adaptive Boosting (AdaBoost) and Support Vector Machines (SVM) together as classifiers. The purpose of this study was to see whether AdaBoost SVM could produce good accuracy with SVM as comparison. Tests were conducted using 10% until 90% data training. Polynomial and RBF kernel were used with number of AdaBoost cycle. The highest accuracy value of SVM was 75% in 90% training data, while the highest accuracy value of AdaBoost SVM was 85.714% in 80% training data. Therefore, it could be that AdaBoost can improve the performance of SVM in classification of OA disease severity.

1. Introduction
Osteoarthritis (OA) is a long-term chronic disease characterized by the deterioration of cartilage in joints which results in bones rubbing together and creating stiffness, pain, and impaired movement [1]. A joint is an area of the body where two different bones are met. OA is the most common chronic disease and joint disability in all ages but usually in elderly people. This disease can affect any joint, but it occurs most often in knees, hips, lower back and neck, small joints of the fingers and the bases of the thumb and big toe. Common risk factors include increasing age, obesity, previous joint injury, overuse of the joint, weak thigh muscles, and genes [2].

In normal joints, a firm and rubbery material which called cartilage covers the end of each bone. Cartilage provides a smooth, gliding surface for the joint motion and acts as a cushion between the bones. In OA, the cartilage breaks down so that causing pain, swelling and problems moving the joint [3]. The severity of OA divided into three categories which are (1) not severe – suspicious narrowing of the joint and potential osteophytes, (2) severe – definitive osteophytes and potential narrowing of the joint gaps on the weight bearing side, and (3) very severe – multiple osteophytes, definitive narrowing of the joint gaps, and potential bone deformities [4].

The progressive development of osteoarthritis is a slow and multi factorial process. The onset of clinical symptoms (pain and progressive loss of function) is usually preceded by molecular and cellular changes in the joint that result in the loss of homeostasis. The diagnosis of osteoarthritis by laboratory tests, X-rays, and a magnetic resonance imaging (MRI) scan [2]. Effective antiosteoarthritic drugs and biological interventions should maintain or restore joint homeostasis and structural integrity. Since the key molecular regulators of these processes remain poorly understood, thus rendering the
development of effectiveness and safe treatments still a major challenge [5]. There is no cure of it, but treatments are available to manage symptoms. Long-term management of the disease will include several factors, managing symptoms included stiffness and swelling, improving joint mobility and flexibility, maintaining a healthy weight, and getting enough of exercise [2].

One way to prevent this disease can be done by doing early detection using machine learning. The detection of OA is important. The earlier treatment can prevent the destruction of cartilage and bone [3]. Machine learning is a branch of science that allows computers to learn based on existing data. Machine learning uses mathematical algorithms implemented as computer programs to identify patterns in large datasets, and to iteratively improve in performing this identification with additional data [6]. One type of this learning is supervised learning for classification. In this study, it was used Adaptive Boosting (AdaBoost) and Support Vector Machines (SVM) together as classifiers.

2. Methods

2.1. Supervised Learning

Supervised learning used targeted training data so that each training data was a pair of data \{(x_1, y_1), \ldots, (x_m, y_m)\}. This learning aimed to form a model that could provide prediction results in accordance with the target for a testing data. Training data was used to construct predictive models to applied into testing data.

Supervised learning could provide discrete predictions called classification. Classification was the arrangement of the system in groups or classes according to established rules or standards.

2.2. Support Vector Machines (SVM)

SVM was one of supervised learning methods developed by Vapnik in 1992 for classification. SVM aimed to resolve the classification problem by forming a hyperplane that maximize the margin.
by dividing the two classes of data [8]. Margin was the closest distance between hyperplane to the nearest point of each class (support vectors). In SVM, the optimal separating hyperplane is determined by giving the largest margin separation \( \rho \) between different class [9].

Given a dataset \( D = \{(x_1, y_1), \ldots, (x_m, y_m)\}, x_t \in \mathbf{X}, y_t \in \mathbf{Y} = \{-1, +1\} \), SVM resolved the following mathematics model:

\[
\min_{w,b} \frac{1}{2} \|w\|^2 \\
\text{s.t.} \quad y_t(w^Tx_t + b) \geq 1, \quad i = 1, \ldots, n
\]  

(1)

For misclassification error cases, it was added \( C \) parameter and slack variable so that SVM mathematics model becomes:

\[
\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \\
\text{s.t.} \quad y_t(w^Tx_t + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, n
\]  

(2)

with \( C > 0 \).

For linearly non-separable cases, it was added kernel function. Kernel function could be defined as

\[
K(x_i, x_j) = \phi(x_i)^T \phi(x_j)
\]  

(3)

There were various kernel functions i.e. [10]:

1) Linear kernel : \( K(x_i, x_j) = x_i^T x_j \)

2) Polynomial kernel : \( K(x_i, x_j) = (x_i^T x_j + 1)^d \)

3) RBF kernel : \( K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \)

2.3. Adaptive Boosting (AdaBoost)

AdaBoost introduced by Yoav Freund and Robert Schapire in 1995. This method aimed to maintain a weight distribution \( \mathbf{w} \) of base classifier (SVM is a base classifier in this paper) iteratively [11]. AdaBoost was an ensemble method that improved the classification results by constructing a set of classifier and combining it. Given a dataset \( D = \{(x_1, y_1), \ldots, (x_m, y_m)\}, x_t \in \mathbf{X}, y_t \in \mathbf{Y} = \{-1, +1\} \), this method performs base classifier training iteratively as many cycles \( t = 1, 2, \ldots, T \). Initial weight vector \( \mathbf{w}^1 \) in this training was arranged the same as follows:

\[
\mathbf{w}_t^i = \frac{1}{m}, \quad i = 1, 2, \ldots, m
\]  

(4)

At each round, the weight vector would be updated until it obtained the right result.

The base classifier worked to find hypothesis \( h_t = \{-1, +1\} \), for \( \mathbf{w}_t \). The quality of hypothesis \( h_t(x_i) \) measured by training error \( \varepsilon_t \) as follows:

\[
\varepsilon_t = \sum_{i=1}^{m} \mathbf{w}_t^i, \quad y_i \neq h_t(x_i)
\]  

(5)

Training error was calculated from the trained weights vector. If \( \varepsilon_t > 0.5 \), then the weighting process was stopped and iteration was not continued.
After the hypothesis $h_t$ accepted, AdaBoost would determine the weight of hypothesis $h_t$, $\alpha_t$. It was obtained $\alpha_t \geq 0$ if $\varepsilon_t \leq 0.5$, and the value of $\alpha_t$ would be increase since $\varepsilon_t$ was decreased. Thus, it was formulated as follows:

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (6)$$

Then, the weight vector $w_t^i$ were updated to

$$w_{t+1}^i = \frac{w_t^i \exp\{-\alpha_t y_i h_t(x_i)\}}{Z_t} = \frac{w_t^i}{Z_t} \times \left\{ \begin{array}{ll} \exp\{-\alpha_t\}, & y_i = h_t(x_i) \\ \exp(\alpha_t), & y_i \neq h_t(x_i) \end{array} \right\}$$

$$\sum_{i=1}^{m} w_{t+1}^i = 1 \quad (7)$$

With $Z_t$ was a normalisation constants that made $\sum_{i=1}^{m} w_{t+1}^i = 1$ so $w_{t+1}^i$ could be distributed.

The hypothesis resulted $H(x)$ based on the number of weights of $T$ hypothesis of the base classifier as follows:

$$H(x) = \text{sign}\left( \sum_{t=1}^{T} \alpha_t h_t(x) \right) \quad (8)$$

This algorithm was based on Schapire & Singer in 1999 [12].

**Table 1.** AdaBoost algorithm

| AdaBoost |
|-------------------------|
| **1. Input:** Dataset $D = \{(x_1, y_1), \ldots, (x_m, y_m)\}$, a Base Classifier algorithm, the number of cycles $T$ |
| **2. Initialize:** the weights of training samples: $w_1^i = 1/m$, for all $i = 1, 2, \ldots, m$. |
| **3. Do** for $t = 1, \ldots, T$ |
| (1) Use the BaseClassifier algorithm to get hypothesis $h_t$, on the weighted training samples. |
| (2) Calculate the training error of $h_t$: $\varepsilon_t = \sum_{i=1}^{m} w_t^i$, $y_i \neq h_t(x_i)$. |
| (3) If $\varepsilon_t > 0.5$; then stop. |
| (4) Set weight for the hypothesis $h_t$: $\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right)$. |
| (5) Update the weights of training samples: $w_{t+1}^i = \frac{w_t^i \exp\{-\alpha_t y_i h_t(x_i)\}}{Z_t}$ $w_t^i \times$ $\left\{ \begin{array}{ll} \exp\{-\alpha_t\}, & y_i = h_t(x_i) \\ \exp(\alpha_t), & y_i \neq h_t(x_i) \end{array} \right\}$, where $Z_t$ is a normalisation constant and $\sum_{i=1}^{m} w_{t+1}^i = 1$. |
| **4. Output:** $H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t h_t(x))$. |

**2.4. AdaBoost SVM**

AdaBoost SVM was AdaBoost method with SVM as the base classifier. Algorithm of this method was similar with algorithm in Table 1. AdaBoost performed the hypothesis weighting of SVM method to obtain a better accuracy. At each cycle, the weight in misclassification error were increased, while the weight on the already well-classified were decreased to reduce the potential weighted back in the next cycle. This process was to do predict the class (label) of hypothesis $h_t$ [11].
Table 2. AdaBoost SVM algorithm.

1. **Input**: Dataset $D = \{(x_1, y_1), \ldots, (x_m, y_m)\}$, SVM algorithm, the number of cycles $T$
2. **Initialize**: the weights of training samples: $w_i^1 = 1/m$, for all $i = 1, 2, \ldots, m$.
3. **Do for** $t = 1, \ldots, T$
   (1) Use SVM algorithm to get hypothesis $h_t$ on the weighted training samples.
   (2) Calculate the training error of $h_t$: $\epsilon_t = \sum_{i=1}^{m} w_i^t, y_i \neq h_t(x_i)$.
   (3) If $\epsilon_t > 0.5$, then stop.
   (4) Set weight for the hypothesis $h_t$: $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$.
   (5) Update the weights of training samples: $w_i^{t+1} = \frac{w_i^t \exp[\{-\alpha_t y_i h_t(x_i)\}]}{Z_t} = \frac{w_i^t}{Z_t} \times \{\exp[-\alpha_t], y_i = h_t(x_i)\}, \{\exp(\alpha_t), y_i \neq h_t(x_i)\}$, where $Z_t$ is a normalization constant and $\sum_{t=1}^{m} w_i^{t+1} = 1$.
4. **Output**: $H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t h_t(x))$.

3. Results

3.1. Used Parameters

Table 3. Used parameters of SVM and AdaBoost.

| Parameter | Value | Explanation |
|-----------|-------|-------------|
| SVM       | Kernel | Polynomial ($d = 2$) | [13], [14] |
|           | RBF   | ($\gamma = 0.01$)   |            |
| AdaBoost  | $T$   | 10           | Based on numerical simulation |

3.2. Numerical Results

The data used for training was OA patient data which was checked by MRI T2Map [15]. Based on those parameters used, the accuracy of SVM result would be compared between SVM with polynomial kernel (polynomial SVM) and SVM with RBF kernel (RBF SVM). The OA data was divided into training data and testing data. We constructed a model using training data and obtained accuracy by testing the model using testing data. Tests were conducted using 10% data training (90% data testing) until 90% data training (10% data testing).

Table 4. Results of OA severity classification using polynomial SVM and RBF SVM.

| Polynomial SVM | RBF SVM |
|----------------|---------|
| **Training Data** | **Accuracy (%)** | **Training Data** | **Accuracy (%)** |
| 10%            | 34      | 10%            | 41,333          |
| 20%            | 31,111  | 20%            | 44              |
| 30%            | 39,250  | 30%            | 41,667          |
| 40%            | 40      | 40%            | 46              |
| 50%            | 49,412  | 50%            | 51,765          |
| 60%            | 57,143  | 60%            | 64,286          |
Based on the result above, it was shown that the highest accuracy value of polynomial SVM was 65% in 90% training data, while the highest accuracy value of RBF SVM was 75% in 90% training data. Thus, RBF SVM accuracy was better than polynomial SVM accuracy.

The accuracy result would be compared between RBF SVM and AdaBoost SVM (with kernel RBF). The purpose of this study was to examine whether AdaBoost SVM could produce better accuracy with SVM as comparison.

| Training Data | Accuracy (%) | Training Data | Accuracy (%) |
|---------------|--------------|---------------|--------------|
| 10%           | 41,333       | 10%           | 53,333       |
| 20%           | 44           | 20%           | 55,556       |
| 30%           | 41,667       | 30%           | 58,333       |
| 40%           | 46           | 40%           | 65           |
| 50%           | 51,765       | 50%           | 70,588       |
| 60%           | 64,286       | 60%           | 71,429       |
| 70%           | 68           | 70%           | 80           |
| 80%           | 68,572       | 80%           | 85,714       |
| 90%           | 75           | 90%           | 80           |

Based on the result above, it was shown that the highest accuracy value of RBF SVM was 75% in 90% training data, while the highest accuracy value of AdaBoost SVM was 85,714% in 80% training data. It was also shown that AdaBoost SVM increased the accuracy of SVM. Thus, it could be concluded that AdaBoost could improve the performance of SVM by maintains the weight of distribution.

4. Conclusion
In this study, AdaBoost SVM were using in the classification of Osteoarthritis (OA) disease severity with SVM as comparison. Based on numerical simulation, the highest accuracy value of SVM was 75% in 90% training data using RBF kernel, while the highest accuracy value of AdaBoost SVM was 85,714% in 80% training data using RBF kernel and number of cycles $T = 10$. Thus, AdaBoost could maintain the distribution weight of SVM iteratively and increasing its accuracy. AdaBoost SVM had better accuracy than SVM in the classification of Osteoarthritis disease severity. For further research, it can be used other base classifier in classifying the disease.

5. References
[1] World Health Organization (WHO)2013 (online) Osteoarthritis ([https://www.who.int/medicines/areas/priority_medicines/BP6_12Osteo.pdf](https://www.who.int/medicines/areas/priority_medicines/BP6_12Osteo.pdf))
[2] Arthritis Foundation 2007 (online) What is Osteoarthritis? ([https://www.arthritis.org/about-arthritis/types/osteoarthritis/what-is-osteoarthritis.php](https://www.arthritis.org/about-arthritis/types/osteoarthritis/what-is-osteoarthritis.php))
[3] Bijlsma JW, Berenbaum F and Lafeber FP 2011 Osteoarthritis: an update with relevance for clinical practice The Lancet 377 pp. 2115–2126
[4] Kellgren J and Lawrence J 2000 Radiological assessment of osteoarthrosis Ann Rheum Dis 16 pp. 494-452
[5] Stampella A, Monteagudo S and Lories R 2018 Wnt signaling as target for the treatment of osteoarthritis Best Practice & Research Clinical Rheumatology xxx pp. 1-9
[6] Lynch CM, Abdollahi B, Fuqua J D, de Carlo A R, Bartholomai J A, Balgemann R N, van Berkel V H, and Frieboes H B 2017 Prediction of Lung Cancer Patient Survival via Supervised Machine Learning Classification Techniques International Journal of Medical Informatics 108 pp. 1-8
[7] Xue H, Yang Q and Chen S 2009 SVM: Support Vector Machines (USA: Taylor & Francis Group)
[8] Panca V and Rustam Z 2017 Application of Machine Learning on Brain Cancer Multiclass Classification The American Institute of Physics (AIP) Conference 1862(1)
[9] Kim H C, Pang S, Je H M, Kim D, and Bang S Y 2003 Constructing support vector machine ensemble Pattern Recognition 36 pp. 2757-2767
[10] Bishop C M 2006 Pattern Recognition and Machine Learning (New York: Springer)
[11] Li X, Wang L and Sung E 2008 AdaBoost with SVM-based Component Classifiers Engineering Applications of Artificial Intelligence 21 pp. 785-795
[12] Schapire R E and Singer Y 1999 Improved Boosting Algorithms Using Confidence-rated Predictions Machine Learning 37 pp. 297-336
[13] Rustam Z and Zahras D 2018 Comparison between Support Vector Machine and Fuzzy C-Means as Classifier for Intrusion Detection System Journal of Physics: Conference Series 1028
[14] Rustam Z and Ariantari N P A A 2018 Support Vector Machines for Classifying Policyholders Satisfactorily in Automobile Insurance Journal of Physics: Conference Series 1028
[15] Pamela C 2017 Korelasi Nilai T2Map dengan Ketebalan Kartilago Lutut Pasien Osteoartritis pada Sekuen Proton Density-Weighted dengan Menggunakan Magnetic Resonance Imaging 1.5 Tesla Tesi Jakarta: Universitas Indonesia

Acknowledgment
This research is funded by Universitas Indonesia via PITTA Research grant 2018 scheme.