

PAPER • OPEN ACCESS

Soft Tissue Tumor Classification using Stochastic Support Vector Machine

To cite this article: Durrabida Zahras *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **546** 052089

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

Soft Tissue Tumor Classification using Stochastic Support Vector Machine

Durrabida Zahras^{1*}, Zuherman Rustam¹, Devvi Sarwinda¹

¹Department of Mathematics, FMIPA Universitas Indonesia, Kampus UI, Depok.

*Corresponding author : durrabida@sci.ui.ac.id

Abstract. As healthcare is becoming one of the most rapidly changing industries by the increasing type of diseases, technology plays an important role in helping medical staffs solve those medical problems. Soft tissue tumors are tumors in the musculoskeletal system that involve soft tissue (tissue other than bone tissue). It includes muscle tissue, nerves, blood vessels, fat, and connective tissue. This soft tissue tumor is divided into two, namely benign and malignant. To prevent any medical errors in classifying patients' data, one machine learning called Stochastic Support Vector Machine is being studied. In this study, we will evaluate soft tissue tumor patients' data in Nur Hidayah Hospital, Yogyakarta, Indonesia using Stochastic Support Vector Machine to see its accuracy. The result is encouraging that Stochastic Support Vector Machine works better than the original Support Vector Machine.

1. Introduction

The World Health Organisation (WHO) has defined health as a state of complete physical, mental and social well-being and not merely the absence of disease or illness or infirmity. We can see the importance of health by how we function as a human being in daily life, without being healthy we cannot do many activities properly. Healthcare is also becoming number one concern of all government around countries. Specifically, tumor care, it sometimes being ignored because it spread slowly. Tumor is simply defined by cells that grows out of control in human body. Soft tissue tumors are the fourth most common malignancy in children, after hematopoietic neoplasm, neural tumor and Wilms tumor [1]. Soft tissue tumors belong to a group of neoplasms that can cause particular problems in their diagnosis and/or treatment. Because they are relatively rare, the individual physician usually has limited experience of these tumors [2].

Because of the development of tumor itself technology plays an important role also they are working hard to improve patients' experiences. In this paper one machine learning is being studied in order to help medical staff solve classifying problem on soft tumor diseases. Machine learning is a branch of artificial intelligence that is closely related to computational statistics, which also focuses on prediction making using computers known as unsupervised learning. The pioneer of machine-learning, Arthur Samuel, defined machine-learning as a "field of study that gives computers the ability to learn without being explicitly programmed." Machine learning focused on classification and regression based on known features previously learned from the training data.

In this paper, one machine learning called Stochastic Support Vector Machine will be studied to classify Soft Tissue Tumor classification problem. We will use the data of Nur Hidayah Hospital Patients and see how this classifier works.



2. Soft Tissue Tumor

Soft tissue can be defined as non-epithelial, extra skeletal tissues of the body exclusive of reticulo-endothelial system, glia and supporting tissues of various parenchymal organs. It is represented by voluntary muscles, fat and fibrous tissue, along with the vessels serving these tissues. By convention, it also includes peripheral nervous system [3].

Soft tissue tumors are defined as mesenchymal proliferations which occur in the extra skeletal non-epithelial tissues of the body, excluding the viscera, coverings of brain and lymphoreticular system. Soft tissue tumors are the fourth most common malignancy in children, after hematopoietic neoplasm, neural tumor and Wilms tumor [3]. The annual incidence of soft tissue tumor is 1.4 per 100000 populations [4]. Soft tissue tumors belong to a group of neoplasms that can cause particular problems in their diagnosis and treatment. Because they are relatively rare, the medical staff usually has limited experience of these tumors. There is a huge variety types and subtypes of soft tissue tumor that grow on human body. That is why blood test does not represent enough that the patients have tumor or not, medical staff's diagnosis is really matters to make decisions.

3. Soft Tissue Tumor Data

The dataset was collected at Nur Hidayah Hospital in Bantul, Yogyakarta, Indonesia. The dataset comprises historic medical record of patients and consultation result made by the doctor about the lump on each patients whether it is called a tumor or not. Table 1 shown the data features. The Medical record data of soft tissue tumor patients represented by 18 factors and 2 target variables: tumor or non-tumor. We will evaluate the data of 76 patients with 2 target variables in Nur Hidayah Hospital.

Table 1. Medical record data attributes.

Number of Attributes	Name of Attributes
1	Age
2	Gender
3	White Blood Count/ Leucocytes (WBC)
4	Red Blood Count/ Erythrocyte (RBC)
5	Hemoglobin (HGB)
6	Hematocrit (HCT)
7	Mean Corpuscular Volume (MCV)
8	Mean Corpuscular Hemoglobin (MCH)
9	Mean Corpuscular Hemoglobin Concentration (MCHC)
10	Platelet Count (PLT)
11	Lymfocytes (%)
12	Monocytes (%)
13	Neutrophils (%)
14	Blood Type
15	Clotting Time (CT)
16	Bleeding Time (BT)
17	Total Protein & Albumin/ Globulin (AGS/AG)
18	Glucose

4. Stochastic Support Vector Machine

This section will explain about how Stochastic Support Vector Machine works. Frequently in practical classification and regression problems, training data cannot be observed precisely because of sampling errors, modeling errors or measurement errors [5]. They are usually presented by random variables. We will investigate the SVM with uncertain input data in a stochastic framework. Then the primal optimization of Stochastic Support Vector Machine written as follow [6]:

$$\min \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \max(0, 1 - \mathcal{Y}_i \mathbf{w}^T \mathbf{x}_i) \right) \quad [1]$$

In stochastic framework, we will apply LaGrange multiplier into the primal optimization and we will use stochastic gradient decent to get the function of $\mathbf{w}$. In general gradient decent, we will minimize the function of $J(\mathbf{w})$ start with the initial guess for $\mathbf{w}$, suppose it is $\mathbf{w}^0$ and we will iterate until it is convergence by computing the gradient of J at $\mathbf{w}^t$ and update $\mathbf{w}^t$ to get $\mathbf{w}^{t+1}$ [6]. Table 2 below shown the algorithm of Stochastic Support Vector Machine.

Table 2. Stochastic Support Vector Machine algorithm [6]

Given training data set $S = \{(x_i, y_i)\}, x \in \mathcal{R}^n, y \in \{-1, 1\}$	
1.	Initialize $\mathbf{w}^0 = 0 \in \mathcal{R}^n$
2.	For epoch = 1 ... T :
1.	Pick a random example (x_i, y_i) from a training set S
2.	Treat (x_i, y_i) as a full dataset and take the derivative of the Support Vector Machine objective as the current $\mathbf{w}^{t-1}$ to be $\nabla J^t(\mathbf{w}^{t-1})$
	$J^t(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \max(0, 1 - \mathcal{Y}_i \mathbf{w}^T \mathbf{x}_i)$
3.	Update: $\mathbf{w}^t \leftarrow \mathbf{w}^{t-1} - \gamma_t \nabla J^t(\mathbf{w}^{t-1})$
3.	Return final $\mathbf{w}$

5. Experiment Results

In this experiment results section, one machine learning called Stochastic Support Vector Machine will be evaluated. We modified the tools of Support Vector Machine in MatLab to suits our data. Correct and incorrect classification results are recorded in a confusion matrix shown in table 3 [7].

TP : Number of samples having cervical cancer which correctly diagnosed
 FP : Number of healthy samples incorrectly identified as having cervical cancer
 TN : Number of healthy samples which are correctly spotted
 FN : Number of samples having cervical cancer incorrectly classified as healthy [7]

$$\text{Accuracy: } \frac{n_{TP} + n_{TN}}{n_{TP} + n_{TN} + n_{FP} + n_{FN}} \times 100\% \quad [2]$$

Table 3. Matrix confusion [7]

Actual Value	Recognize Value	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Shown on table 4, the experiment results using Stochastic Support Vector Machine using kernel RBF and sigma 0,05. There are four aspects considered as a result: Accuracy, Sensitivity, Precession, and Fscore. Accuracy explained how the data being classified, it means the data that classified into its

rightful classes, that is explained its accuracy. The second one is the sensitivity; sensitivity is a measure of the proportion of actual positive cases that got predicted as positive or true positive. Sensitivity is also called as Recall. This implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative. Mathematically, sensitivity can be calculated as the following:

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad [3]$$

Precision is the positive predictive value or the fraction of the positive predictions that are actually positive. It can be calculated as the following:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad [4]$$

Fscore is used when we want to seek the balance between precision and recall. The formula is:

$$\text{Fscore} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad [5]$$

Now, take a look at table 4, the highest accuracy which is 71.43% goes to when we take 90% data training and the lowest accuracy is 54.55% goes to when we take 70% data training while the average of the sensitivity reach 69.21%.

Table 4. Results on Soft Tissue Tumor Data using Stochastic Support Vector Machine into 2 classes.

Data Training (%)	Accuracy (%)	Sensitivity	Precision	Fscore	Sigma	Kernel	Computational Time
10	70.15	69.23	1.00	1.971522453	0.05	RBF	0.02
20	73.33	71.43	1.00	1.972386588	0.05	RBF	0.00
30	61.54	67.44	0.05	0.092146863	0.05	RBF	0.03
40	66.67	69.23	0.08	0.164940896	0.05	RBF	0.03
50	62.16	67.74	0.05	0.099689065	0.05	RBF	0.03
60	63.33	66.67	0.08	0.164933339	0.05	RBF	0.08
70	54.55	64.71	0.03	0.053505849	0.05	RBF	0.08
80	60.00	75.00	0.01	0.029550827	0.05	RBF	0.05
90	71.43	71.43	1.00	1.972386588	0.05	RBF	0.16
Average	64.80	69.21	0.37	0.72	-	-	0.05

We also classify the soft tissue tumor data with the original Support Vector Machine without adding the stochastic gradient decent to the function of $\mathbf{w}$. The result shown on table 5 explained that the accuracy of Stochastic Support Machine is higher than Support Vector Machine. The accuracy of Support Vector Machine only reach 57.82% while Stochastic Vector Machine reach 64.80%.

Table 5. Results on Soft Tissue Tumor Data using Support Vector Machine into 2 classes.

Data Training (%)	Accuracy (%)	Sensitivity	Precision	Fscore	Sigma	Kernel	Computational Time
10	31.34	60.71	0	0.66	0.05	RBF	0.02
20	66.67	66.67	1.00	1.97	0.05	RBF	0.02
30	61.54	67.44	0.04	0.092	0.05	RBF	0.02
40	60.00	65.79	0.04	0.095	0.05	RBF	0.08
50	67.57	67.57	1	1.97	0.05	RBF	0.03
60	63.33	65.52	0.15	0.31	0.05	RBF	0.03
70	31.82	60.71	0	0.66	0.05	RBF	0.03
80	66.67	72.73	0.03	0.07	0.05	RBF	0.06
90	71.43	80.00	0.03	0.07	0.05	RBF	0.02
Average	57.82	60.71	0.26	0.66	-	-	0.03

6. Conclusion

To conclude, one machine learning is being studied which is Stochastic Support Vector Machine. It generally works on finding the best hyperplane that can separate the data into two classes, Hyperplane is a linear pattern whose maximum margin gives the maximum separation between the decision classes [8]. As we can see on table 4, the use of Stochastic Support Machine is quite good shown on the average of its accuracy is higher compare to the original Support Vector Machine. It means that each data given were classified into its rightful classes. The Accuracy of Stochastic Support Vector Machine reach 64,80% while the original Support Vector Machine only reach 57.82%.

As a result, we can summarize that Stochastic Support Vector Machine can be a good classifier to classify the soft tissue tumor patient data. Along with this study, we hope that it can help all medical staff solve some classifying problem. For future works, this research can be continued to find a better method to solve cervical cancer risk data classification problem.

7. Acknowledgement

This research was financially supported by University of Indonesia, with PIT.9 2019 research grant scheme (ID Number NKB-0039/UN2.R3.1/HKP.05.00/2019)

References

- [1] Rosenberg AE. Bones, Joints and Soft Tissue Tumors. Chapter-26 In: Robbins and Cotran Pathologic Basis of Disease by Kumar, Abbas and Fausto, 7th Edition, Saunders: Elsevier, 2004; 1273-1324.
- [2] K. Kathrin, K. Detlef. Review Article Soft Tissue Tumors “New Perspectives on Classification and Diagnosis”. Deutsches Ärzteblatt International Dtsch Arztebl Int 2009; 106(39): 632–6
- [3] Weiss SW, Goldblum JR. General Considerations. Chapter-1 In: Enzinger and Weiss’s Soft Tissue Tumors. 4th Edition, St. Louis: Mosby, 2001: 1-19.
- [4] Rydholm A, Berg NO. Epidemiology of Soft tissue sarcoma in locomotor system. A retrospective population based study of interrelationships betweenclinical and morphological variables. Acta Pathol Microbiol Immunol Scand. 1984; 92A: 363.
- [5] A. Maryam, E. Sohrab, Stochastic Support Vector Machine for Classifying and Regression of Random Variables, Neural Process Lett (2018) 48:1–29 <https://doi.org/10.1007/s11063-017-9697-0>
- [6] University of Utah Outline. Support Vector Machines: Training with Stochastic Gradient Descent. Machine Learning. Fall 2017. Machine Learning. Spring 2018.
- [7] Normed kernel function-based fuzzy possibilistic C-means (NKFPCM) algorithm for high-dimensional breast cancer database classification with feature selection is based on Laplacian

Score, Lestari. A. W, Rustam Z. Conference Paper in AIP Conference Proceedings 1862(1):030143 • July 2017 DOI: 10.1063/1.4991247 Conference: Conference: International Symposium on Current Progress in Mathematics and Sciences 2016 (Iscpms 2016): Proceedings of the 2nd International Symposium on Current Progress in Mathematics and Sciences 2016

- [8] Rustam, Z., & Zahras, D. (2018, June). Comparison between Support Vector Machine and Fuzzy C-Means as classifier for intrusion detection system. In Journal of Physics: Conference Series (Vol. 1028, No. 1, p. 012227). IOP Publishing.