Performance Analysis of Support Vector Machine in Sex Classification of The Sacrum Bone in Forensic Anthropology

Iis Afrianty¹, Dewi Nasien², Habibollah Haron³

¹Department of Informatic Engineering, Faculty of Science and Technology, ²Department of Informatic Engineering, Faculty of Science Computer, ³School of Computing, Faculty of Engineering

¹Universitas Islam Negeri Sultan Syarif Kasim Riau, ²Institut Bisnis dan Teknologi Pelita Pekanbaru, ³University of Technology Malaysia

¹Jl. H.R. Soeberantas km 11,5 Simpang Baru Panam, Pekanbaru, Riau, ²Jl. Jend. Ahmad Yani No.78-88, Pekanbaru, ³Sultan Ibrahim Chancellery Building, Jl. Iman, 81310 Skudai, Johor, Malaysia

E-mail: ¹iis.afrianty@uin-suska.ac.id, ²dewinasien@lecturer.pelitaindonesia.ac.id, ³habib@utm.my

Article: Accepted: April 26, 2022
Revised: May 27, 2022
Issued: June 15, 2022

© 2022 The Author(s).

This is an open access article under the CC BY-SA license

*Correspondence Address:
iis.afrianty@uin-suska.ac.id

ABSTRACT

Sex classification is part of forensic anthropological identification aimed at determining whether the skeleton belongs to a male or a female. This paper exhibits the performance of the Support Vector Machine (SVM) in classifying the sex of the sacrum in forensic anthropology. Bone data was measured by the metric method based on six variables, namely superior breadth, anterior length, mid ventral breadth, real height, diameter the base, and max-transverse diameter of the base. This study shows performance analysis of SVM using the library libSVM with linear, polynomial, and RBF kernel to observe the results of the comparison of the accuracy of the kernel used. According to the results of the trials, the best accuracy was attained in each kernel function, i.e., the RBF kernel is 83.33% with γ = 1 and C = 1, the polynomial is 85.56% at γ = 2, C = 2 and d =1, and the linear kernel obtained best accuracy is 84.44 % with C = 2 and C = 3. In conformity with the experimental result, polynomial attained the highest accuracy of 85.56% at γ = 2, C = 2, and d =1.

Keywords: Forensic Anthropology; Sacrum Bones; Sex Classification; Support Vector Machine;
I. INTRODUCTION

Forensic anthropology is the discipline of skeletal identification, particularly with regard to the recovery and analysis of skeletons aimed at knowing the biological profile [1]. From previous studies, identification of skeletal remains can be done using Deoxyribose Nucleic Acid (DNA) analysis on laboratory or radiological examinations. The most popular method of determining the sex of the found bones is using DNA analysis [2]. However, in some cases, if the skeletal remains are in a burnt, dismembered (not intact) or very dry condition, DNA analysis fails to provide accurate results because DNA proteins cannot be extracted under the conditions stated. Thus, important parameters present in the bone cannot provide information [2, 3]. Based on the disadvantages of DNA analysis in this condition, forensic anthropology was developed to improve the overall identification of important parameters. This aims to obtain a more reliable characterization of individuals and provide more data to confirm the identity of the biological profile, particularly in determining sex.

Forensic anthropology can identify skeletons that are burnt, dry, or not intact to recognize important biological profiles (Figure 1), including the cause and time of death [1]. Forensic anthropology has four important parameters in identification of skeleton or skeletal remains, namely sex, age, ancestry, and stature in which sex determination is a major step in identifying the human biological profile [1, 4-6].

Sex classification is aimed at determining whether a given skeleton belongs to a male or female [3]. Knowledge about the sex of the body of an unknown collection is very important to make a more accurate estimate of the age [3]. Without an accurate sex determination, it will not be possible to accurately estimate the age at death. Thus, sex determination is necessary for further identifying the age, ancestry, and stature estimations [2].

In determining sex, bone data were analyzed using metric or morphological measurement methods [7]. Metric measurements relate to sizes (such as weight), body proportions as well as the shape of the human skeleton (such as pubic angle, pubic length). While the morphological measurement is related to the observation of visual criteria [3]. Some parts of the body skeleton that are usually analyzed in determining sex are the pelvic [8-11], skull [12-15], mandible [16], cranial [17], femur [18-21], and tibia [22].

The pelvic bones are known as a reliable and best part of the skeleton to reach a diagnosis in sex determination [7, 18]. The pelvic bones consist of a pair of hip bones, the coccyx, and the sacrum. The pelvic is another element of the skeleton that exhibits sexual differences. Both metrically and morphologically, the female pelvic is wider than the male [7]. The sacrum bones are a part of the pelvic bones that are closely related to reproduction and fertility [3]. Therefore, if all the required sacrum bone data is complete, it can be used as a more accurate sex determination indicator up to 100% [3, 7]. In this study, the data of the sacrum was used with measurement of the metric method which became the variable in determining sex.

In determining sex in previous studies, various classification techniques have been applied to determine sex with statistical certainty measures [22, 23]. Classification technique is part of the supervised learning approach. Supervised learning is divided into classification and regression depending on the output. In classification, the training data is
labeled in such a way that the input and the desired output correspond to each other. Classification technique is a technique used to solve the classification problem.

The application of classification techniques has been widely applied in forensic anthropology, particularly in the case of sex determination. In previous studies, the most popular classification techniques in determining sex were Discriminant Function Analysis (DFA) which was studied by [10, 15, 22, 24, 25] and Logistic Regression (LR) by [18, 25-27]. In contrast, over the last few years there has been a trend in forensic anthropology to adopt and apply machine learning (ML) approaches [11, 23].

ML approaches are becoming a trend in determining the sex of skeletal remains, as research conducted by [2, 10, 11, 14, 17, 28, 29]. ML is a branch of computer science that has the ability to learn and predict future outcomes with invisible data [11, 28]. The most common ML technique used in sex determination, such as Support Vector Machine (SVM) [2, 5, 23, 30, 31].

SVM is consistently used in various classification studies because it is known to provide high accuracy, strong generalization, and high stability [32]. SVM is a binary statistical classification method that can optimally separate two classes. SVM model is a ML method based on statistical learning theory with a focus on minimizing structural risk [33].

In previous studies, SVM was used to determine the sex of the bones, most of which came from bone image data. This paper uses metric measurements and the application of SVM in classifying the sex of the sacrum bone which is intended to test the performance of SVM based on the test parameters of each kernel.

II. METHODOLOGY

This study has several steps, and the first step is a literature review conducted relating to the classification of sex. The second step is collecting data, The third step is pre-processing the data. The next step is to divide the data using cross-validation models, then process using the SVM method and evaluate the model performance using a confusion matrix.

2.1 Literature review of data measurement

In the sex classification process, there are two methods of measuring data, namely morphologic and metric methods. The morphologic method is the observation of sexual traits on bones. It has the advantage of obtaining results quickly with high classification accuracy if the bone is available and the observer has sufficient experience [3]. On the other hand, the metric method is based on measurements and statistical analysis [26]. Metric focused on linear measurements, indices, or angle measurements that primarily captured the size differences between females and males [22]. Metric measurements are preferred because they have high accuracy, are easy to perform, and do not require any special skills [26]. Hence this metric measurement is superior to morphology and can be evaluated quantitatively [7, 26]. In this study, metric measurements were used to analyze the sacrum bone, to obtain more accurate results.

2.2 Data collection

The sacrum is the part of the pelvic bone that is involved with reproductive and fertility functions. The data used in this paper included 91 sacrum bones (34 Females and 57 Males) derived from analysis of previous studies, namely from [3], which used the BPNN method.

There are six measurements for sacrum bones used as indicators variable in determining sex, namely real height, mid-ventral breadth, superior anterior breadth, anterior length, the anterior-posterior diameter of the base, and max-transverse diameter of the base. Figure 2 shows the measurement of the sacrum bones.

![Figure 2. The measurement of the sacrum bone is based on six variables [3]](image)

Table 1 and Table 2 below show the measurement variables of the sacrum bone with their respective codes.
Table 1. Six variables for sacrum bone measurement

<table>
<thead>
<tr>
<th>Code</th>
<th>Variables in the sacrum bones</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH</td>
<td>Real Height</td>
</tr>
<tr>
<td>MB</td>
<td>Mid-ventral Breadth</td>
</tr>
<tr>
<td>AL</td>
<td>Anterior Length</td>
</tr>
<tr>
<td>ASB</td>
<td>Anterior Superior Breadth</td>
</tr>
<tr>
<td>APDB</td>
<td>Anterior Posterior Diameter of the Base</td>
</tr>
<tr>
<td>TDB</td>
<td>Max-Transverse Diameter of the Base</td>
</tr>
</tbody>
</table>

Table 2. Sacrum bone measurement (mm)

<table>
<thead>
<tr>
<th>Code</th>
<th>RH</th>
<th>AL</th>
<th>ASB</th>
<th>MB</th>
<th>APDB</th>
<th>TDB</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>120</td>
<td>99</td>
<td>80</td>
<td>21</td>
<td>43</td>
<td>45</td>
<td>F</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>95</td>
<td>108</td>
<td>118</td>
<td>84</td>
<td>29</td>
<td>45</td>
<td>45</td>
<td>M</td>
</tr>
</tbody>
</table>

2.3 Data pre-processing

The data that has been measured according to the variable indicators (Table 2), is then pre-processed the data. Data pre-processing is an important step to achieve good classification performance before evaluating data on machine learning [34]. In this paper using the transformation process. The data transformation carried out is to change the sex class into a number (class for female = 1 and male = 0), then use the normalization process to change the value of the variable so that it has a range of values that are not too far apart. Data normalization is an important pre-processing step that involves transforming features within the same range [34]. Normalization is a good way to reduce data discrepancy. There are three methods of data normalization, namely, z-score, min-max, and decimal scaling [34]. The normalization used is the min-max normalization, which changes the scale of each variable to an interval of 0.0 to 1.0 or -1 to +1 by computing equation (1) below [34].

\[
\nu' = \frac{\nu - \text{min}_A}{\text{max}_A - \text{min}_A} \left(\text{new}_{\text{max}_A} - \text{new}_{\text{min}_A}\right) + \text{new}_{\text{min}_A}
\]

(1)

The advantage of min-max normalization is that it maintains all data value relationships exactly. The results of the transformation calculations can be seen in Table 3 below.

Table 3. Transformed data values

<table>
<thead>
<tr>
<th>Code</th>
<th>RH</th>
<th>AL</th>
<th>ASB</th>
<th>MB</th>
<th>APDB</th>
<th>TDB</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH</td>
<td>0.280</td>
<td>0.755</td>
<td>0.194</td>
<td>0.394</td>
<td>0.000</td>
<td>0.222</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>RH</td>
<td>0.280</td>
<td>0.510</td>
<td>0.806</td>
<td>0.515</td>
<td>0.571</td>
<td>0.333</td>
<td>0</td>
</tr>
</tbody>
</table>

After the transformation (normalization) process is carried out, then the bone data is divided using 10-cross validation where the data is divided into 10 equal-sized partitions. Nine partitions are used for the training process, while another partition is used for the classification testing process. After the data is divided, it is continued in the training and testing process using the Support Vector Machine (SVM).

2.4 Support Vector Machine (SVM)

SVM is one of the most popular methods in the two-class classification in Machine Learning (ML) technique. ML techniques are divided into three categories, namely supervised learning, unsupervised learning, and reinforcement learning [35]. SVM aims to find a linear separating hyperplane maximizing the distance to the nearest individuals of each of the two classes, called margins [2]. SVM can be used as a classification or regression algorithm [36, 37]. The original idea of the SVM method is to issue two classes, one above the first class vector and the other below the second class vector [38].

SVM algorithms have a good application to issue two classes and provide excellent classification performance [32, 39]. SVM constructs a set of hyperplanes that separates data into categories [36]. SVM issues a hyperplane (linear boundary) for the two data classes [40]. In classification problems, particularly regarding gender classification, class is retrospectively assigned as a label +1 for male and -1 for female outcomes. The SVM classification method have kernel parameter, namely Linear, Sigmoid, RBF, and Polynomial. The RBF kernel is often used because it provides fairly accurate classification results [41].

This paper uses LibSVM as a classifier of the pelvic bones. LibSVM is a kernel-based software library that uses multiclass SVM. The classification in LibSVM was performed using linear, polynomial, and RBF kernels with 10-fold cross-validation. Appropriate kernel
functions can reduce the number of calculations, and the experimental data increases the recognition rate up to a certain limit.

The selection of kernel functions is key to achieving an accurate SVM classification [39]. The Confusion Matrix method was used to measure SVM performance with accuracy parameters. There are several commonly used kernel functions, namely RBF, linear, and in this paper the following kernel functions are used:

- Linear kernel: \( K(x_i, x_j) = x_i^T \cdot x_j \)  \( \text{(2)} \)
- Polynomial kernel: \( K(x_i, x_j) = (\gamma x_i^T \cdot x_j + r)^d \)  \( \text{(3)} \)
- RBF: \( K(x_i, x_j) = e^{-(\gamma ||x_i-x_j||^2)} \), \( \gamma > 0 \)  \( \text{(4)} \)

where \( \gamma, d, r \) are kernel parameters for each kernel function.

III. RESULTS AND DISCUSSION

The experiment was carried out using 91 sacrum bones with six variables that have been implemented using the Rapid Miner. Sacrum bone data is divided using 10 cross-validations in which the data is divided into 10 equal-sized partitions. Nine partitions are used for the training process, while one more partition is used for the classification testing. Experiments were executed by selecting three different kernel functions: RBF, polynomial, and linear kernel. The model accuracy is measured using the Confusion Matrix. The accuracy results obtained from the different SVM kernels can be seen in the following Table 4-8.

Linear SVM Kernel is a good kernel function when the data is separated linearly. The parameters used are \( C = 1, 2, \) and 3. The accuracy value obtained can be seen in Table 4.

### Table 4. Comparison of the accuracy (%) by Linear kernel functions

<table>
<thead>
<tr>
<th>Linear</th>
<th>Accuracy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( C = 1 )</td>
<td>82,22</td>
<td></td>
</tr>
<tr>
<td>( C = 2 )</td>
<td>84,44</td>
<td></td>
</tr>
<tr>
<td>( C = 3 )</td>
<td>84,44</td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Comparison of the accuracy (%) by RBF kernel functions

<table>
<thead>
<tr>
<th>RBF</th>
<th>( \gamma = 1 )</th>
<th>( \gamma = 2 )</th>
<th>( \gamma = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C = 1 )</td>
<td>83,33</td>
<td>82,22</td>
<td>81,11</td>
</tr>
<tr>
<td>( C = 2 )</td>
<td>82,22</td>
<td>81,11</td>
<td>81,11</td>
</tr>
<tr>
<td>( C = 3 )</td>
<td>82,22</td>
<td>81,11</td>
<td>78,89</td>
</tr>
</tbody>
</table>

The polynomial kernel SVM is based on a similar approach to the linear kernel. In the kernel polynomials depend not only on one particular feature of the input sample, but also on their combination in determining their similarity [42]. Comparison of the accuracy results obtained by polynomial kernel with \( \gamma = 1, 2, \) and 3 can be seen in the following Table 6-8.

### Table 6. Comparison of the accuracy (%) by Polynomial kernel functions with \( \gamma = 1 \)

<table>
<thead>
<tr>
<th>( \gamma = 1 )</th>
<th>( d = 1 )</th>
<th>( d = 2 )</th>
<th>( d = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C = 1 )</td>
<td>82,22</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>( C = 2 )</td>
<td>84,44</td>
<td>80</td>
<td>79</td>
</tr>
<tr>
<td>( C = 3 )</td>
<td>84,44</td>
<td>80</td>
<td>80</td>
</tr>
</tbody>
</table>

### Table 7. Comparison of the accuracy (%) by Polynomial kernel functions with \( \gamma = 2 \)

<table>
<thead>
<tr>
<th>( \gamma = 2 )</th>
<th>( d = 1 )</th>
<th>( d = 2 )</th>
<th>( d = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C = 1 )</td>
<td>84,44</td>
<td>80</td>
<td>76,78</td>
</tr>
<tr>
<td>( C = 2 )</td>
<td>85,56</td>
<td>78,89</td>
<td>74,56</td>
</tr>
<tr>
<td>( C = 3 )</td>
<td>84,44</td>
<td>78,89</td>
<td>72,33</td>
</tr>
</tbody>
</table>

### Table 8. Comparison of the accuracy (%) by Polynomial kernel functions with \( \gamma = 3 \)

<table>
<thead>
<tr>
<th>( \gamma = 3 )</th>
<th>( d = 1 )</th>
<th>( d = 2 )</th>
<th>( d = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C = 1 )</td>
<td>84,44</td>
<td>78,89</td>
<td>72,33</td>
</tr>
<tr>
<td>( C = 2 )</td>
<td>84,44</td>
<td>77,78</td>
<td>70,22</td>
</tr>
<tr>
<td>( C = 3 )</td>
<td>83,33</td>
<td>77,78</td>
<td>68</td>
</tr>
</tbody>
</table>

Pertaining to the results of the experiments that have been executed, the best accuracy obtained in each kernel function, i.e., the RBF kernel is 83.33% with \( \gamma = 1 \) and \( C = 1 \), the polynomial is 85.56% at \( \gamma = 2 \), \( C = 2 \) and \( d = 1 \), and the linear kernel obtained best accuracy is 84.44% with \( C = 2 \) and \( C = 3 \). Figure 2 contains the overall accuracy obtained for a
The proposed model with different kernel functions for SVM.

![Accuracy](image)

**Figure 3.** Comparison of the accuracy results of the proposed SVM kernel function

Thus, it can be seen that the selection of kernel functions and kernel parameters can affect the level of accuracy.

**IV. CONCLUSION**

Classification approaches have been widely used in various areas, including forensic anthropology. This study uses sacrum bones for sex classification. Data is partitioned using ten cross-validations, which divides the data into ten equal-sized segments. The training phase makes use of nine divisions, whereas the classification testing procedure makes use of one extra partition. Experiments were conducted using RBF, polynomial, and linear kernel using three distinct kernel functions. The Confusion Matrix is used to assess model correctness.

From the following data presentation process, the experiment results revealed the best accuracy in each kernel function. The RBF kernel is 83.33% with $\gamma = 1$ and $C = 1$, the polynomial is 85.56% at $\gamma = 2$, $C = 2$, and $d = 1$, and the linear kernel obtained best accuracy is 84.44% with $C = 2$ and $C = 3$. Within the result of 85.56% found, this study is in line with other previous findings by [4], [10], which reported high accuracy more than 80%.

Sex classification of the sacrum bones using Support Vector Machine (SVM) in forensic anthropology has been conducted and implemented in this work. In applying SVM to sex classification using LibSVM tools. In this study, SVM kernel parameters have been compared, namely RBF, Polynomial, and Linear Kernel. The experimental result shows that the kernel function affects the recognition rate. Likewise, the kernel function parameters can affect the level of accuracy.

In this study, the polynomial kernel obtained the highest accuracy compared to linear and RBF kernel, 85.56% at $\gamma = 2$, $C = 2$, and $d = 1$. Meanwhile, research conducted by [31, 37, 43] states that the use of the RBF kernel has higher accuracy than other kernels. The kernel function influences this evidence in SVM; in other words, the effectiveness of SVM depends on the selection of kernel functions and parameters of each kernel. Besides, this study still has shortcomings in terms of the number of bone samples used.

In future work, sex classification from sacrum bones can be improved by applying the other techniques specific to improve the accuracy and analysis performance of classification techniques such as the Hidden Markov Model, and the optimization of SVM kernel parameters using Particle Swarm Optimization. In addition, it can also be improved by adding the number of pelvic-bones data and calculating the processing time of the techniques used.

**BIBLIOGRAPHY**


and Control, Studies in Computational Intelligence 575, 2015. Springer International Publishing Switzerland 2015, doi: https://doi.org/10.1007/978-3-319-11017-2_11


I. Afrianty, etc: Performance analysis of...


I. Afrianty, etc: Performance analysis of...
