Snake bite classification using Chain code and K nearest neighbour

Rayiemas Manggala Putra¹, Adiwijaya² and Dody Qori Utama³
¹,²,³ School of Computing, Telkom University, Bandung

E-mail: masrayiemas@student.telkomuniversity.ac.id¹, adiwijaya@telkomuniversity.ac.id², dodyqori@telkomuniversity.ac.id³

Abstract. Live in tropical country causes a high risk of Indonesians affected by snake bites. One of the causes of death due to snake bites is incorrect snake bite identification by visually identifying. The difference between a venomous and non-venomous snake on their anatomy resulting in different bite marks on the victim. Bite marks are useful for medical team for identifying venomous snake bites or not, but there is no system that helps this problem. This research tries to build a system to recognize the existing bite points on the snake bite image and then classified to venomous snake bite or non-venomous using Chain Code and K Nearest Neighbour. The result of recognizing the bite marks has sensitivity value of 75.95%, specificity of 52.7% and accuracy of 65%. Classification of snake bites categories has sensitivity value of 76.92%, specificity value of 85.71% and accuracy of 80%.

1. Introduction
Live in tropical country causes a high risk of Indonesians affected by snake bites [1]. Snake bite occur when the snake feels threatened, although the victim did not intend to hurt even more if they want to harm or kill the snake [2]. Yet an estimated 5.4-5.5 million people are bitten by snakes each year, resulting in about 20,000 and 125,000 causes death [3]. In 2007, Indonesia have 12,739 - 214,883 snake bite case estimated with between 20 and 11,581 deaths [4]. Some causes of death due to snake bites are errors in giving first aid, only treated traditionally, never managed to reach the hospital to get help and incorrect snake bite identification by visually identifying [4, 5].

Bite marks are useful for medical team for identifying venomous snake bites or not [6], but can be done well by someone who knows how to identify snake bites correctly. There is no information technology-based systems that handle this to help ordinary people who do not know about the differences of venomous and non venomous snake bite. For that reason, a system has been built to help the detection of bite marks and their position then classification into venomous or non venomous snake bite category by using Chain Code method for feature extraction and K Nearest Neighbour for classification. Chain Code method was chosen based on research previously conducted by Sun, Zhang, Liu, Zhu on Shape Feature Extraction of Fruit Image Based on Chain Code [7]. K Nearest Neighbour (KNN) is selected as the classifier because according to Bhatia [8], they explained about advantages using KNN are very fast training, simple and easy to learn and robust to noisy training data.

Some of limitations to this research were the use of a data set of 20 images consists of 13 venomous snake bite and 7 non venomous snake bite with uneven image quality. Image obtained
from Dr. dr. Tri Maharani, M.Sc., SP.EM., WHO handbook [9, 10] and internet [11, 12, 13, 14, 15, 16, 17]. All images have been cut in the wound area with a size of 400x400 pixels. The difficulty in obtaining data on snake bite wound in this study causes authors taking data from internet, but with strong data information. The focus of this research is the system can recognize bite points and non bite points then classification into venomous or non venomous bites.

2. Related works
In this world there are two categories of snake namely venomous and non venomous. Both of these categories have differences in the structure of their teeth. The structure of the venomous snake’s teeth has two fangs that are useful for injecting venom to their prey, while non venomous snakes do not have these two fangs on their teeth structure [2]. Hubbard [18] wrote on his book if there are 2 until 4 bite points on venomous snake bite mark, while non venomous snake leave bite marks according to the pattern of their teeth structure.

There is no research yet on the classification of venomous snake bites based on the bite mark using image processing technology. The research to detect venomous snake bite or not largely through laboratory tests like Kalana’s research. Kalana [19] researched to detect the venomous snake bite or not by a simple blood test using Phospholispase A₂ enzyme which is commonly contained in snake venom, so this test detects the high activity of Phospholispase A₂ enzyme in the victim’s blood to find out snake venom in the victim’s body.

However in 1995, Nishioka [6] in their research Bite Marks are Useful for The Differential Diagnosis of Snakebite in Brazil they researched 42 images of snake bites cases then they got 89% recognize venomous snake bites and 100% recognize non venomous snake bite by identifying the bite mark images visually, but can be done well by someone who knows how to identify snake bites correctly.

Sun et al. [7] proposed shape feature extraction using chain code method and convert the chain code value to some feature like circumference, area, shape complexity, shape compactness, inscribed circle radius, circular degree, concave rate, height and width of object. The linear regression of height/width ratio between data extracted using chain codes and from manual measurement shows high precision.

The latest research in 2017 was conducted by James [5] about the classification of snake species based on images using image processing technology. He used 31 features which are the taxonomic part of the snake’s body with an approach using the Nearest Neighbor method. The results of F1 Score obtained for the classification of snake species using 31 features of the taxonomic part of the snake are 95.9%.

3. Proposed scheme
Data train used in this paper based on extraction result of 9 venomous snake bite images and 5 non venomous snake bite images are 158 points with the proportion of data 81 bite points and 77 non bite points. Data test used in this paper based on extraction result of 4 venomous snake bite images and 2 non venomous snake bite images are 46 points with the proportion of data 32 bite points and 14 non bite points. Figure 1 is venomous and non venomous snake bites image.

![Figure 1. Snake bites image.](image)
This research will build a system to identify snake bite points or non bite points in the image and then the image is categorized as a bite of a venomous snake or not. Generally, the system is divided into several parts including the preprocessing, followed by feature extraction at each point or object in the image, then the bite point recognition process and finally the classification process into the venomous snake bite category or not as shown in the figure 2.

![Figure 2. System flowchart.](image)

### 3.1. Preprocessing

All image data used before entering the feature extraction stage will pass the preprocessing stage first. The block diagram of the preprocessing stage is illustrated in figure 3.

![Figure 3. Preprocessing block diagram.](image)
(a) Image data is filtered to produce smoother images. The filter used is Median Filter with 7 kernel.

(b) Grayscaling is the process of changing the color of the image to gray so that the image matrix becomes 2-dimensional.

(c) The gray colored image is then finalized which changes the pixel value of the image to 0 or 1 only. This process uses the imbinarize function with adaptive thresholds in MATLAB programming. The parameters used are ForegroundPolarity with a dark value and a sensitivity value of 0.5.

(d) The image binarization results are then filtered a second time to minimize noise. The filter used remains Median Filter with 7 kernel.

(e) Image complement changes image value 0 to 1 and vice versa. This process will make it easier for objects to be identified with a value of 1 and the background with a value of 0.

(f) Erosion is an image processing technique to attenuate the area around pixels or objects that value is 1, while Dilation is an image processing technique to thicken the area around a pixel or an object that value is 1. If these two operations are used sequentially, it is called the Opening technique which is useful for separating several objects that are quite close together [20].

(g) The removal of pixels is to minimize noise and objects that are non bite points. The deleted object is an object that has pixels less than 50 and more than 1400 pixels.

(h) Objects that can affect the search for other bite points are objects that are connected to the outer boundary of the image, then any object attached to the outer boundary of the image will be deleted.

(i) All remaining objects will be taken one by one according to each pixel size.

(j) Objects that have been cut according to each pixel will be generalized to a size of 25x25 pixels.

3.2. Feature extraction
Basically the chain code method has a simple concept that is to explore the edge (edge) of an object, but in the process produces a code for each direction that passes. Freeman [21], first introduced the concept of chain code method which describes the sequence of pixel boundaries using the concept of 8-connectivity or 4-connectivity, but in this study the 8-connectivity method is used so that the direction of each movement is an encode by using a scheme numbering \{i| i = 0, 1, 2, \cdots, 7\} that anticlockwise at an angle of 45 degrees for each i. The coding process starts from the far left pixel point of the object then follows the object’s shape by rotating counterclockwise until it returns to the starting pixel point. Each time move to another pixel, the direction of displacement is represented by the code in accordance with the 8-connectivity concept as shown in figure 4. The results of the chain code are in the form of data arrays where each element is an encode value that represents the edge of an object in the image [22]. The process of coding the edges of objects can be described as in figure 5.
According to Sun [7] research, value of chain code that have been obtained converted to some value of feature such as circumference, area, shape complexity, shape compactness, circular degree of object. Circumference calculation is using formula:

$$L = \sum_{i=1}^{N} n_i$$

(1)

This equation explained that the circumference of an object is the length of the pattern of the object’s edge. N is the total number of codes and this is the value that is in the code i in the eight-way chain code where the even code has a value of 1 while the odd code has a value of $\sqrt{2}$ [7]. Chain code can be represented in the Cartesian plane in the form of Cartesian coordinates, the relationship of the chain code value with the Cartesian plane is shown in table 1.

Table 1. Relation between chain code value to Cartesian plane.

| Chain Code | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------|---|---|---|---|---|---|---|---|
| X axis ($a_j$) | 1 | 1 | 0 | -1 | -1 | 0 | 1 |
| Y axis ($a_k$) | 0 | 1 | 1 | 1 | 0 | -1 | -1 |

To get the coordinate points based on chain code value by using Equation (2). The abscissa ($x$) and ordinate ($y$) values for code $i$ are the sum of the previous abscissa and ordinate points with the value of $a_j$ and $a_k$ for the code $i$. Initial value of abscissa and ordinate are respectively 0 [23]. Furthermore, by using these coordinate points, can be calculated the regions value by using Equation (3).

$$S = \frac{1}{2} \sum_{i=1}^{k} (x_iy_{i+1} - x_{i+1}y_i)$$

(3)

Regions S is absolute value where $k$ is the sum of all coordinates that form objects from the starting point to the coordinate point before the starting point and in the equation when $i = k$ then $x_{k+1} = x_1$ and $y_{k+1} = y_1$ [24].

$$e = \frac{L^2}{S}$$

(4)

$$F = \|L^3\| / (4\pi S)$$

(5)

$$R_0 = 4\pi S/L^2$$

(6)

The next features are obtained by using the circumference and area values that has been obtained previously. Shape complexity can be reflected with the index value $e$ in equation (4).
In equation (4), the value of $e$ represents the shape of the perimeter size per unit area. The greater the value of $e$, the larger the graph of the circumference size in the unit area and the graph becomes more complicated. Furthermore, the shape compactness feature can be calculated as $F$ which describes the graph of object compactness in equation (5). Circular degree ($R_0$) describes the degree of an object approaching roundness in equation (6). The range of $R_0$ values is 0 to 1. The greater the value of $R_0$, the more circular objects. For each equation above the $L$ value is circumference and $S$ is the area [7].

3.3. Classification

The classification method used to identify the bite point in the image is K Nearest Neighbor (KNN). This method works by identifying data classes that are not yet known based on the data of the nearest neighbor whose class is known [25, 5]. Data that will be identified is calculated distance metrics of features for each data. The distance metrics technique used in this study is Euclidean Distance with the formula for calculating the distance shown in equation (7):

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$$ (7)

The distance value that has been obtained is used as a reference value for the ascending sorting process. Next, choose the top data of a number of K that has been predetermined the guess class is determined by the majority of the class in the K sample data [25].

The classification of images into the category of venomous snake bites or not by using the results of the bite point recognition and then calculated the number of identified bite points. Snake bite wound categorized as venomous snake bite if the number of bites is 1 to 4 bite points, more than that the image of snake bite wound is categorized as a non-venomous snake bite [18].

4. Results and discussion

4.1. System evaluating with 5-fold

The first scenario is to test the performance of system by using a 5-Fold technique, which means all data (data test and data train) totaling 204 points are divided into 5 sections. Each section will alternately be a test data, while the rest will be data train and the system will run 5 times. All data previously is randomly positioned, but in same random position for each time the system run with a different $K$ value for K Nearest Neighbour. $K$ value used in this test is an odd number from 1 to 20.

| K   | Bite Points Recognition | Image Category Classification |
|-----|-------------------------|-------------------------------|
|     | Sensitivity  | Specificity | Accuracy | Sensitivity | Specificity | Accuracy |
| 1   | 62.62%      | 56.06%      | 59.32%   | 69.23%      | 71.43%      | 70.00%   |
| 3   | 71.69%      | 50.78%      | 62.14%   | 53.85%      | 71.43%      | 60.00%   |
| 5   | 72.82%      | 50.59%      | 62.05%   | 46.15%      | 71.43%      | 55.00%   |
| 7   | 71.62%      | 48.43%      | 60.55%   | 53.85%      | 85.71%      | 65.00%   |
| 9   | 74.27%      | 50.26%      | 63.05%   | 53.85%      | 85.71%      | 65.00%   |
| 11  | 70.62%      | 48.01%      | 60.14%   | 61.54%      | 71.43%      | 65.00%   |
| 13  | 75.95%      | 52.70%      | **65.00%** | 76.92%      | 85.71%      | **80.00%** |
| 15  | 76.46%      | 45.01%      | 60.00%   | 61.54%      | 100.00%     | 75.00%   |
| 17  | 79.11%      | 47.02%      | 64.45%   | 53.85%      | 100.00%     | 70.00%   |
| 19  | 82.12%      | 44.47%      | 65.00%   | 46.15%      | 100.00%     | 65.00%   |

Table 2. Results of System Evaluating with 5-fold.
Table 2 shows that the best result of sensitivity, specificity and accuracy obtained by using K value equal 13 with sensitivity average value of 75.95%, specificity average value of 52.7% and average accuracy value of 65%, it means this system can recognize the bite points about 75.95%, the non bite points about 52.7% and the system can distinguish the bite points and non bite points about 65%. In classification to the venomous bite mark or not this system obtained the best sensitivity value of 76.92%, specificity value of 85.71% and accuracy value of 80%, it means 10 of 13 venomous snake bite image can be classified correctly and 6 of 7 non venomous snake bite image can be classified correctly. If the classification image results associated with the results of recognizing the bite points especially for venomous snake bite shows if there is 1 to 3 miss classification of bite points and there is at least 1 miss classification of non bite points.

4.2. System evaluating with data test to data train
The second scenario is to test the performance of system using the data test and data train by previously separating 6 images consisting of 4 images of poisonous snake bites and 2 images of non-venomous snake bites as data test and 14 images consisting of 9 images of poisonous snake bites and 5 images of non-venomous snake bites as data train. K for K-Nearest Neighbor used in this test is an odd value number from 1 to 20.

| K  | Bite Points Recognition | Image Category Classification |
|----|------------------------|------------------------------|
|    | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | Accuracy |
| 1  | 34.38%      | 71.43%      | 45.65%    | 100.00%     | 50.00%      | 80.00%    |
| 3  | 56.25%      | 64.29%      | 58.70%    | 100.00%     | 100.00%     | 100.00%   |
| 5  | 75.00%      | 50.00%      | 67.39%    | 100.00%     | 100.00%     | 100.00%   |
| 7  | 78.13%      | 71.43%      | **76.09%** | 100.00%     | 100.00%     | **100.00%** |
| 9  | 68.75%      | 71.43%      | 69.57%    | 100.00%     | 100.00%     | 100.00%   |
| 11 | 75.00%      | 64.29%      | 71.43%    | 100.00%     | 100.00%     | 100.00%   |
| 13 | 71.88%      | 71.43%      | 71.43%    | 100.00%     | 100.00%     | 100.00%   |
| 15 | 78.13%      | 64.29%      | 73.91%    | 100.00%     | 100.00%     | 100.00%   |
| 17 | 78.13%      | 57.14%      | 71.74%    | 100.00%     | 100.00%     | 100.00%   |
| 19 | 78.13%      | 57.14%      | 71.74%    | 100.00%     | 100.00%     | 100.00%   |

Table 3 shows that the best result of sensitivity, specificity and accuracy obtained by using K value equal 7 with sensitivity value of 78.13%, specificity value of 71.43% and accuracy value of 76.09%. It means this system can recognize the bite points about 78.13%, the non bite points about 71.43% and this system can differentiate the bite points and non bite points about 76.09%. In classification to the venomous bite mark or not this system obtained the best sensitivity, specificity and accuracy value is 100%, it means the system can classify 4 venomous snake bite images and 2 non venomous snake bite images correctly. This result is obtained because the majority of images in the data test have good conditions, such as adequate lighting, good resolution and the wound in the image is new enough so little noise or other objects are detected and have similarities with the actual bite points.

4.3. Find the best K to reach optimum result of each data test
The third scenario is to run the system for each image in the data test to all data train used in the second scenario to determine which K value is needed so that the bite point and not the bite on each data test image can be identified optimally, because of various data quality can allow each data to achieve optimum results with different K values. K for K-Nearest Neighbor used in this test is an odd value number from 1 to 20.
Figure 6 until figure 11 are the graphs of sensitivity, specificity and accuracy for each image in the data test. Figure 6 shows that image 1 has the most optimum sensitivity, specificity and accuracy values of 100%, 100% and 100% using K values of 1 and 3. Figure 7 shows that image 2 has the most optimum sensitivity, specificity and accuracy values of 92.31%, 100% and 92.86% using K values of 7. Figure 8 shows that image 3 has the most optimum sensitivity, specificity and accuracy values of 100%, 100% and 100% using K values of 7 until 19. Figure 9 shows that image 4 has the most optimum sensitivity, specificity and accuracy values of 54.55%, 66.67% and 58.82% using K values of 11 until 15. Figure 10 shows that image 5 has the most optimum sensitivity, specificity and accuracy values of 100%, 100% and 100% using K values of 3 and 7. Figure 11 shows that image 6 has the most optimum sensitivity, specificity and accuracy values of 100%, 0% and 100% using K values of 13 until 19. From all figures above, it shows that every data has their own K value to reach their optimum value of sensitivity, specificity, accuracy and some data has different K value compared by the best K value on second scenario.
5. Conclusions
Based on the results of several testing scenarios performed it can be concluded that overall system can recognize bite points and non bite points using Chain Code as shape feature extraction of bite points and K Nearest Neighbour as classifier with sensitivity value of 75.95%, specificity value of 52.7% and accuracy of 65%. Then the system can categorize the image into venomous snake bite or not, based on recognize the bite points with sensitivity value of 76.92%, specificity value of 85.71% and accuracy of 80%. These results are obtained with the best K value of 13.

The system has been very good at recognizing bite points, but it is still not good enough at recognizing non bites points. This is proven by the specificity value that is smaller than the sensitivity value, meaning that there are still many non bite points that are recognized as bite points, which results in incorrect classification of images. However, if the system processes data with good quality such as in terms of lighting, resolution or condition of bite wounds it will minimize another objects are detected and system can distinguish the bite points and non bite points very well then it can categorize images correctly.

In this research, preprocessing was still not optimal due to uneven data quality. It is expected that further research will do different preprocessing techniques and have evenly data quality. Addition of feature selection techniques can also be considered. Subsequent research can also be developed to calculate the distance between bite points that can be useful to identify snake species of the bite mark.

References
[1] Anjana Silva et al. “Awareness and perceptions on prevention, first aid and treatment of snakebites among Sri Lankan farmers: a knowledge practice mismatch?” In: Journal of Occupational Medicine and Toxicology 9.1 (2014), p. 20.
[2] Barry S Gold, Richard C Dart, and Robert A Barish. “Bites of venomous snakes”. In: New England Journal of Medicine 347.5 (2002), pp. 347–356.
[3] Anuradhani Kasturiiratne et al. “The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths”. In: PLoS medicine 5.11 (2008), e218.
[4] Randy Adiwinata and Erni J Nelwan. “Snakebite in Indonesia”. In: Acta Medica Indonesiana 47.4 (2015).
[5] Alex James. “Snake classification from images”. In: PeerJ Preprints 5 (2017), e2867v1.
[6] Sérgio De A Nishioka, Paulo Vitor P Silveira, and Fernando A Bauab. “Bite marks are useful for the differential diagnosis of snakebite in Brazil”. In: Wilderness & environmental medicine 6.2 (1995), pp. 183–188.
[7] Yong-Xianga Sun et al. “Shape feature extraction of fruit image based on chain code”. In: Wavelet Analysis and Pattern Recognition, 2007. ICWAPR’07. International Conference on. Vol. 3. IEEE, 2007, pp. 1346–1349.
[8] Nitin Bhatia et al. “Survey of nearest neighbor techniques”. In: arXiv preprint arXiv:1007.0085 (2010).
[9] David A Warrell et al. “Guidelines for the management of snake-bites.” In: Guidelines for the management of snake-bites. (2010).
[10] David A Warrell et al. “Guidelines for the management of snake-bites, 2nd edition.” In: Guidelines for the management of snake-bites. (2016).
[11] Rottalma. Snake bites and picture. http://www.reptileforums.co.uk/forums/snakes/969098-snake-bites-pictures.html. [Online; accessed 22-Jun-2018]. 2013.
[12] Donna Fernstrom. Do python snake bites hurt bad on humans. https://www.quora.com/Do-python-snake-bites-hurt-bad-on-humans. [Online; accessed 22-Jun-2018]. 2017.
[13] Gibby. *Snake bites and picture*. http://www.reptileforums.co.uk/forums/snakes/326960-most-painful-snake-bite-comparing-11.html. [Online; accessed 23-Jun-2018]. 2009.

[14] Jesse Dorsett. *Snake Bite*. http://www.abc.net.au/news/2017-01-24/snake-bite/8208208. [Online; accessed 17-Jul-2018]. 2017.

[15] Jorji Harper. *Jorji Harper’s tiger snake bite wound*. http://www.abc.net.au/news/2015-11-02/jorji-harper-tiger-snake-bite-wound/6906472. [Online; accessed 17-Jul-2018]. 2015.

[16] IndianSnakes.org. *Presentation “Snake Bite Facts First-aid and Prevention”*. http://indiansnakes.org. [Online; accessed 17-Jul-2018]. 2015.

[17] Caroline Fountain. *Man bitten by cottonmouth snake in Huntersville shares story*. http://www.fox46charlotte.com/news/local-news/man-shares-his-story-after-bit-by-cottonmouth-while-fishing-in-huntersville. [Online; accessed 17-Jul-2018]. 2017.

[18] J. Hubbard. *The Survival Doctor’s Complete Handbook*. Reader’s Digest, 2016. Chap. 5, pp. 90–91. ISBN: 9781621453062.

[19] Kalana Maduwage, Margaret A O’leary, and Geoffrey K Isbister. “Diagnosis of snake envenomation using a simple phospholipase A2 assay”. In: *Scientific reports* 4 (2014).

[20] Rafael C. Gonzalez. *Digital Image Processing Using Matlab*. 2nd ed. MHE, 2011, pp. 500–501. ISBN: 978-0070702622.

[21] Ming-Kuei Hu. “Visual pattern recognition by moment invariants”. In: *IRE transactions on information theory* 8.2 (1962), pp. 179–187.

[22] Mingqiang Yang, Kidiyo Kpalma, and Joseph Ronsin. *A survey of shape feature extraction techniques*. 2008.

[23] Ernesto Bribiesca. “A new chain code”. In: *Pattern Recognition* 32.2 (1999), pp. 235–251.

[24] Daniel Zwilinger. *CRC standard mathematical tables and formulae*. 30th ed. Chapman and Hall/CRC, 1996, pp. 500–501.

[25] Thomas Cover and Peter Hart. “Nearest neighbor pattern classification”. In: *IEEE transactions on information theory* 13.1 (1967), pp. 21–27.