Sheet C1 Pattern Number Recognition Using Artificial Immune System Optimization

Yuniarti Denita Sari[1], Tri Basuki Kurniawan[1*], Edi Surya Negara[1], Tata Sutabri[2]

[1] Magister Teknik Informatika, Universitas Bina Darma, Palembang, Sumatera Selatan
[2] Ilmu Komputer, Universitas Bina Darma, Palembang, Sumatera Selatan

[*]**Email**: tribasukikurniawan@binadarma.ac.id

Abstract

When the general election is held in Indonesia, Sheet C1 serves as a record of the results of vote counting at each polling location (TPS). Human error makes it difficult to enter accurate vote data during the recapitulation phase, where the number of votes is a crucial piece of information for the conduct of general elections. The problem was that the KPPS (Voting Committee Group) failed to clearly write the legitimate vote acquisition for each TPS on sheet C1, which caused difficulty for the KPU employees when entering the data. In order to recognize the numerical patterns on sheet C1, we, therefore, require a method that uses the Artificial Immune System (AIS), specifically the CSA (Clonal Selection Algorithm) algorithm, which is one of the AIS system's algorithms for optimization. The CSA method will be used in this research to accomplish number pattern recognition and to examine the operation of number pattern character recognition.

Keywords

C1 Sheet; Pattern Recognition; Artificial Immune System; Clonal Selection Algorithm; Machine Learning

Introduction

The results of the vote count at each polling place (TPS) during the general election are recorded on Sheet C1. Up until the general election's final results are known, the vote count results will be recapitulated in stages. Human error causes problems with valid vote acquisition data supplied during the recapitulation process, hence it is thought that a solution is required to account for this. This is necessary for election outcomes that involve the collection of valid votes, as the total number of votes is crucial information for the conduct of general elections. The problem KPU officers had entering the data was a result of the KPPS of each TPS not clearly putting down the votes that were legitimate on page C1. As a result, one of the areas of machine learning that focuses on the technique of classifying things into specific classes in order to solve certain issues, number pattern recognition, is required. There are various problems that can be resolved with pattern recognition, including handwritten character recognition, voice recognition, human facial identification systems, DNA and fingerprint pattern recognition, and the recognition of spam or non-spam e-mails.

Identification systems and signature image pattern recognition utilizing an Artificial Neural Networks system with the Backpropagation method (Zaitun, Warsito, & Pauzi, 2015), as well as an analysis study of Arabic (Indian) handwriting pattern recognition using the K-Nearest approach, have all been solved using pattern recognition. Neighbors and Connected Component Labeling (Akbar & Sarwoko, 2016), Principal Component Analysis (Ardiansyah, 2013), Neocognitron (Rosnelly, 2018), Jawi Handwriting Character Recognition (Fikri, Arnia, & Muharar, 2016), Handwriting Number Pattern Recognition (Astried & Kurniawan, 2009), Neighbors and Connected Component Labeling (Huri, Nazri, Bakar, & Kurniawan, 2011), and Handwriting Pattern Recognition on Presidential and Vice-Presidential Voting Forms (Yang, Li, Hu, Wang, & Zou, 2014). Therefore, we require a technique that utilizes Artificial Immune Systems to recognize the pattern of numbers on sheet C1 (AIS).

It is hoped that Artificial Immune Systems (AIS), which are inspired by immunological theory and the observed functions, principles, and models of immunity, will be able to cope with changes brought on by various disorders. AIS is a system that is able to adapt itself to changes that occur in the surrounding environment. problem. Additionally, AIS has been used to solve issues in the fields of DNA design using a single objective artificial immune system (Huri et al., 2011), A Survey of Artificial Immune System Based Intrusion Detection (Yang et al., 2014), An Immune Clonal Selection Algorithm for Synthetic Signature Generation (Song & Sun, 2014), A Clone Selection Based Real-Valued Negative Select (Zhang & Xiao, 2018), and Applying the Artificial Immune System algorithm to increase process speed in the color ordering and mapping method with an eight-neighborhood approach (Robandi, 2019).

In this study, 3534-character data were gathered from sheet C1 utilizing 589 samples from each TPS in the Penukal Abab Lematang Ilir (PALI) district. The character data taken were the character data of candidate pairings (paslon) 1 and 2. Due to the fact that each character data for candidate pair 1 and 2 is made up of three characters, there are six-character data to be processed on each sheet of C1, for a total of 3534-character data that come from 589 sheets of C1 multiplied by the six character data on each sheet of C1.

Considering the aforementioned context of the issue, the researcher is drawn to the topic Recognition of Number Patterns on Sheet C1 Using Optimization of the Artificial Immune System (Artificial Immune System).

Methodology

Sheet C1

The results of the vote count records kept by KPPS (Voting Organizing Group) 3 and 4 at each TPS (Polling Station) during the general election are shown on Sheet C1. The Republic of Indonesia's KPU recapitulates votes in stages, starting at the sub-district level and moving up to the regency/municipal, provincial, and final KPU levels. The results will be scanned during the recapitulation procedure in this district or city. Where is the C1 model form scan result? Anyone may view, access, and download the Plano from the SITUNG (Voice Counting Information System). The goal of making a copy of this C1 document public is to increase accountability and

transparency in the way elections are conducted, as well as to motivate more people to get involved and work together to realize the people's sovereignty or a strong state. It also aims to reduce the opportunity for fraud during any stage of the voting process.

Artificial Immune System (AIS)

An example of artificial intelligence that mimics the functioning of a biological thing's immune system is the "Artificial Immune System" (humans). AIS mimics the body's antibody system, which the body uses to defend against pathogens (external objects or disturbances). A further benefit of AIS is that it learns more quickly and stores information more effectively (memorizing) to enable the immune system to respond more quickly when a pathogen enters for the first time. to combat the pathogen. The reaction required to defeat the disease is also quicker if there are pathogens of the same sort and degree present. This occurs as a result of immune system cells binding to and storing strategies for dealing with certain infections. Every disease that affects the body will continue this scenario. The immune system can be sped up to respond to viruses by learning and memorization, preventing the immune system from first analyzing pathogens that enter the body. Simply put, the immune system must adapt so that the infections that enter the body line up with their knowledge (Robandi, 2019).

Currently, AIS has found widespread use in a number of industries, including computer security, scheduling, anomaly approach, failure detection, optimization, and disturbance recovery planning. The four approaches that AIS owns are the Dendritic Cell Algorithm, Immune Network Algorithm, Clonal Selection Algorithm, and Negative Selection Algorithm. Clonal Selection Algorithm will, however, be the technique employed in this study.

Clonal Selection Algorithm (CSA) for Optimization

The basic reaction of the adaptive immune system is described by the clonal selection algorithm (Robandi, 2019). This algorithm uses the immune system's mechanism to help choose cells for cloning depending on how well they can recognize pathogen antigens (foreign objects from outside the body). In order to determine which cells will be cloned or reproduced and which cells will be left to die in order to increase the ability to identify and bind antigens, this algorithm also includes a positive selection process and a negative selection process.
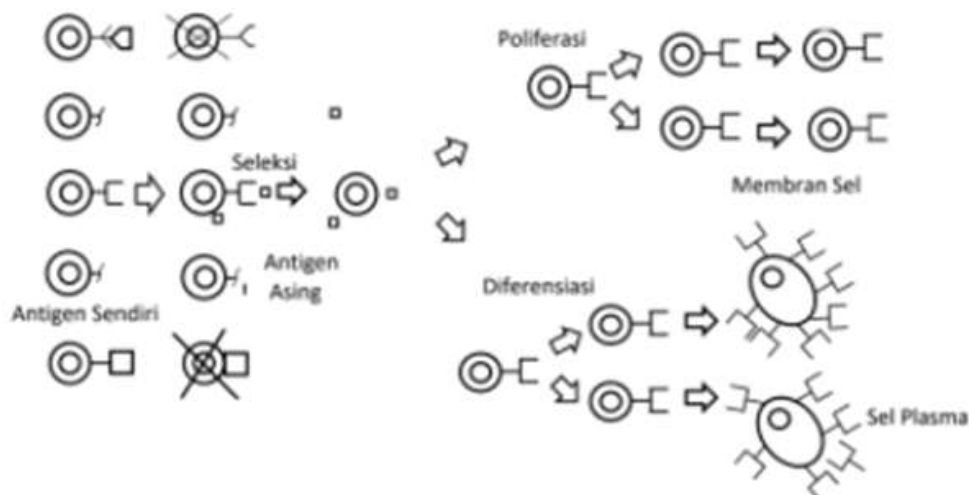

Figure 1. Work on Clonal Selection

The image at figure 1 above is an example of how a cell might proliferate (cloning). Cells that are able to detect their antigens can multiply, but cells that are unable to do so cannot. Clonal selection will occur in cells that can identify their own cell antigens (CS). In situ clonal selection's key characteristics include:

1. New cells are clones of their parents and are prone to a mechanism that causes a high rate of mutation (somantic hypermutation).
2. Removal of newly formed lymphocytes with self-reactive receptors, which are tiny leukocytes with primary functions in the immune system.
3. Adult cells' proliferative and differentiating responses to antigens (Robandi, 2019).
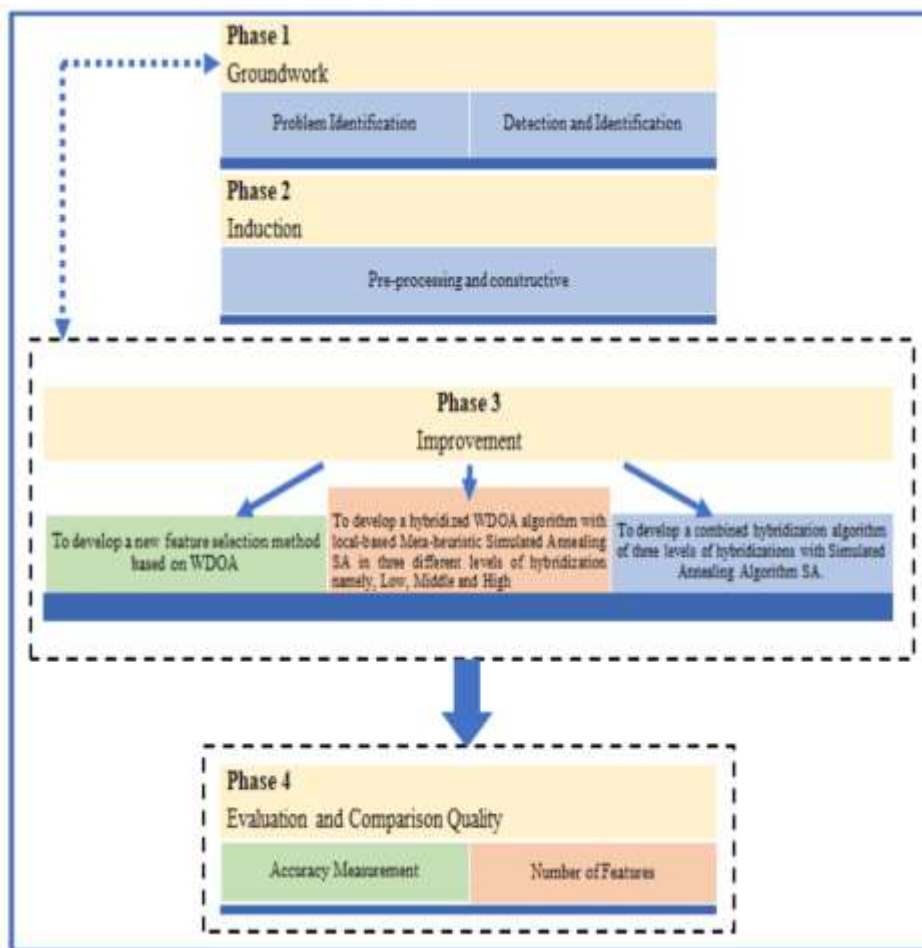

Figure 2. The Research Framework of Text Classification

Problem Identification

The goal of this task is to identify the relevant related studies. We are mainly focusing on understanding the issues facing the development of an effective text classifier. This task was realized by reviewing the most recent text classifiers and their details to identify the existing

techniques' strengths and drawbacks. The methodology of text classification is illustrated in figure 2.

Two major text classifiers- related problems were identified in this phase. The identified problems are, first, the ambiguity performance of three classifiers with the aim of detecting the class of new documents. Second, feature reduction while maintaining high performance in FS problems.

Detection and Identification Cases

The advancement of the Internet and the increased amount of online information has significantly impacted the ability to detect and identify huge documents. The classification approach is capable of simulating human thinking, and it has been successfully applied in various areas due to its salient features in mining. However, text classification to detect and identify new events, documents, or sentiments is complicated. This is because a large volume of data degrades classifiers' performance due to the high dimensionality of feature space.

Results and Discussion

Data Analysis

The research on number pattern recognition on sheet C1 begins with the preparation phase, which entails conducting data analysis. The sample of the C1-PPWP Sheet is depicted in figure 3. At this stage, the preparation is carried out by researchers making initial observations or direct observations of objects, specifically by looking for information on election results, such as the general election for president and vice president in 2019. The total valid votes from each candidate pair or pair of candidates published on the C1-PPWP sheet on the KPU official website, referred to as SITUNG (Voice Counting Information System), are the calculation results that become the subject of study. The researcher will demonstrate one of the obtained C1-PPWP forms in the paragraphs that follow.

With a total of 589 polling stations (TPS) scattered throughout 5 (five) Districts, the Penungkal Abab Lematang Ilir (PALI) area in South Sumatra Province served as the study's primary data source.

    a. Abab, there are 8 Kelurahan with a total of 82 TPS.
    b. There are 13 Kelurahan in Penungkal, totaling 92 TPS.
    c. North Penungkal, which consists of 13 sub-districts and 74 TPS altogether.
    d. Talang Ubi, which consists of 20 sub-districts and 243 TPS.
    e. Tanah Abang, which consists of 17 sub-districts and has 98 TPS altogether.

Figure 3. C1-PPWP Sheet

Data Selection

The researcher will choose the data that will be used as the test sample at the data selection stage, which is conducted following the data analysis stage. The sample data used are the results of the valid votes cast for each pair of presidential and vice-presidential candidates (paslon) from each sheet or sheet C1-PPWP. This sheet or sheet C1-PPWP focuses on data taken from the Penungkal Abab Lematang Ilir (PALI) area in South Sumatra Province with a population of approximately 1.2 million people. There are 5 (five) Districts totaling 589 Total Polling Places (TPS) for the general election.

Each presidential and vice-presidential candidate consists of three characters, making a total of 6 characters of data on sheet C1-PPWP, which is the subject of the research to be processed in the study. This is the total number of legitimate votes for candidate pairs 1 and 2. The following figure 4 displays the required character data.

Figure 4. Data on valid votes for the presidential and vice-presidential candidates

In order to acquire 6 characters of handwritten pattern character data, the data that the author has circled for each number will next be cropped; an example of the cropped handwriting pattern is shown in Figure 5.

Figure 5. Cropped Handwriting Pattern

The 6-character data will be multiplied by 589 C1 sheets, making a total of 3534 character data that needs to be processed. Then, each character's chopped data will be saved as a jpeg file with different pixel sizes. Each jpg file has the file name format "X.URUTAN," which stands for the sequence number of each file that has been cropped to match the numeric character (0-9) or letter x written on the C1-PPWP sheet and "X," which stands for the type of character according to numbers (0-9) or the letter x on the sheet. The cropped handwriting pattern will be categorized based on the class of each handwriting form once all the character data has been removed. 11 classes—0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and x—are produced from the classification that was done.

Data Set
Before being processed, the sample data is categorized by first grouping each image that was obtained based on each number. The sample data is in the form of an image that has been cropped. so that this is how the data set folder layout looks:

Figure 6. Data Set for Number Figure

Preprocessing

In order to ensure that every cropped image has the same pixel size before data processing and testing are done, each existing dataset is preprocessed at this point by being scaled to 15x21 pixels. It can be seen in figure 7 below.

```
LENGTH_IMAGE  = 15
HIGH_IMAGE = 21
ARRAY_SIZE = LENGTH_IMAGE * HIGH_IMAGE  # Features number


Mapping = {"0": 0, "1": 1, "2": 2, "3": 3, "4": 4, "5": 5, "6": 6, "7": 7, "8": 8, "9": 9, "X": 10}
reverseMapping = {0: "0", 1: "1", 2: "2", 3: "3", 4: "4", 5: "5", 6: "6", 7: "7", 8: "8", 9: "9", 10: "X"}

img_dir = ('dataset/')
categories = ['0','1','2','3','4','5','6','7','8','9', 'X']

data = []

print('Reading data is starting process')
for category in categories:
    path = os.path.join(img_dir, category)
    label = Mapping[category]

    idx = 1
    for img in os.listdir(path):
        imgpath = os.path.join(path, img)
        image = cv2.imread(imgpath,0)
        image = cv2.resize(image,(LENGTH_IMAGE, HIGH_IMAGE))

        print(imgpath)
        plt.imshow(image)
        plt.show()

        image = np.array(image).flatten()
```

Figure 7. Preprocessing Dataset

Classification

The data set is split into training data and testing data during the classification stage. It can be seen in figure 8 below.

```python
print("Splitting data training dan testing ...")
X_train, X_test, Y_train, Y_test = \
    sklearn.model_selection.train_test_split(number, label, test_size=0.1, random_state=5, shuffle=True)
print("Data Training: [", len(Y_train), "] data")
print("Data Testing: [", len(Y_test), "] data")
print("Splitting data Done!")
print()

train_set = []
for idx in range(len(X_train)):
    x = X_train[idx]
    y = Y_train[idx]
    train_set.append((x, y))

test_set = []
for idx in range(len(X_test)):
    x = X_test[idx]
    y = Y_test[idx]
    test_set.append((x, y))

ais = AIS(hyper_clonal_rate=20, clonal_rate=0.8, class_number=11, mc_init_rate=0.4,
          total_num_resources=10, affinity_threshold_scalar=0.8, k=3, mutation_rate = 0.2, max_iter = 3)

print('Training Processing ...')
accuration = ais.train(train_set)
print('Training Done!')
print()
```

Figure 8. Dataset Classification

Process of the Clonal Selection Algorithm (CSA)

The Artificial Immune System (AIS) method and Clonal Selection Algorithm (CSA) technology are combined in the final stage to begin testing for the detection of all data sets that have previously been uploaded to Jupiter Notebook as numerical pictures as shown in figure 9.

```
AB[_class].append(ARB(vector=mc_match.vector, _class=mc_match._class, mutation_rate=self.MUTATION_RATE))
stim = mc_match.stimulate(antigene)

iterations = 0
while True:
    iterations += 1
    MAX_CLONES = int(self.HYPER_CLONAL_RATE * self.CLONAL_RATE * stim)
    num_clones = 0
    while num_clones < MAX_CLONES:
        clone, mutated = mc_match.mutate()

        if mutated:
            AB[_class].append(clone)
            num_clones += 1

    # Competition for resources
    avgStim = sum([x.stimulate(antigene) for x in AB[_class]]) / len(AB[_class])

    MIN_STIM = 1.0
    MAX_STIM = 0.1

    for c in AB.keys():
        for ab in AB.get(c):
            stim = ab.stimulate(antigene)
            if stim < MIN_STIM:
                MIN_STIM = stim
            if stim > MAX_STIM:
                MAX_STIM = stim
```

Figure 9. Application of the Clonal Selection Algorithm (CSA)

Based on the findings of the testing procedure conducted using the Clonal Selection Algorithm (CSA) approach and the Artificial Immune System (AIS) method, the accuracy results were 61.82%, as shown in Figure 10, and the confusion matrix results are presented in Figure 11 below.

```
Training time : 5773.4819 seconds
Training Done!

Testing Processing ...
Testing time : 206.4310 seconds
Testing Done!

Accuracy : 61.82 %

Predict Processing ... [ 110 ] data
Predict time : 205.8217 seconds
Prediction Result:
[4, 7, 7, 7, 6, 7, 4, 9, 7, 7, 7, 4, 3, 4, 10, 2, 7, 6, 8, 2, 10, 6, 7, 8, 1, 4, 5, 4, 0, 6, 5, 8, 1, 7, 7, 10, 0, 0,
9, 0, 3, 9, 0, 9, 1, 2, 7, 3, 5, 9, 8, 7, 10, 8, 6, 3, 4, 10, 4, 8, 3, 6, 9, 2, 1, 10, 7, 7, 7, 7, 0, 0, 6, 7, 10, 8,
0, 7, 2, 6, 6, 2, 2, 3, 6, 8, 5, 2, 0, 2, 10, 0, 7, 8, 6, 7, 0, 6, 2, 6, 10, 4, 7, 2]
Predict Done!
```

Figure 10. Results of the testing process

```
Confusion Matrix:

[[11, 0, 0, 0, 0, 0, 2, 1, 0, 0, 0],
 [0, 4, 0, 0, 0, 0, 0, 5, 0, 0, 0],
 [0, 0, 6, 0, 0, 0, 0, 1, 1, 0, 0],
 [0, 0, 1, 4, 0, 0, 0, 0, 0, 1, 0],
 [0, 0, 0, 1, 11, 0, 0, 1, 0, 1, 1],
 [0, 0, 1, 0, 0, 5, 2, 1, 0, 2, 0],
 [0, 0, 0, 0, 0, 0, 5, 1, 4, 0, 0],
 [0, 0, 0, 0, 0, 0, 2, 8, 0, 0, 0],
 [0, 0, 3, 0, 0, 0, 2, 1, 3, 0, 0],
 [1, 0, 0, 1, 0, 0, 0, 0, 1, 3, 0],
 [0, 0, 0, 0, 0, 0, 0, 4, 0, 0, 8]]

Confusion Matrix Done!
```

Figure 11. Results of the Clonal Selection Algorithm (CSA) Technique in the Confusion Matrix

According to the findings of the research, it can be determined that the pattern of numbers on sheet C1 can be detected using the Artificial Immune System (AIS) approach and the Clonal Selection Algorithm (CSA) technology. Additionally, it can be noted that the results of number pattern recognition and the accuracy are 61.82% based on tests conducted on all number picture patterns that are processed using the Python programming language and one of the tools in the Anaconda application, namely Jupiter Notebook.

Conclusion

It may be concluded that the Artificial Immune System (AIS) method and the Clonal Selection Algorithm (CSA) technology can be used to detect the pattern of numbers on sheet C1. Additionally, it should be highlighted that testing on all number picture patterns processed with the Python programming language and one of the tools in the Anaconda application, namely Jupiter Notebook, yielded results for number pattern identification and accuracy of 61.82%.

References

Akbar, R., & Sarwoko, E. A. (2016). Studi Analisis Pengenalan Pola Tulisan Tangan Angka Arabic (Indian) menggunakan Metode K- Nearest Neighbors dan Connected Component Labeling. *Dinamika Rekayasa*, *12*(2), 45. https://doi.org/10.20884/1.dr.2016.12.2.144

Ardiansyah, R. F. (2013). Pengenalan Pola Tanda Tangan Dengan Menggunakan Metode Principal Component Analysis (PCA). *Fakultas Ilmu Komputer Universitas Dian Nuswantoro*, *2*, 14 pages.

Astried, A., & Kurniawan, T. B. (2009). Peningkatan Kecepatan Proses Pada Metode Color Ordering Dan Mapping Dengan Pendekatan Delapan-Ketetanggaan. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *7*(1), 51. https://doi.org/10.12928/telkomnika.v7i1.576

Fikri, R., Arnia, F., & Muharar, R. (2016). Pengenalan Karakter Tulisan Tangan Jawi

Menggunakan Metode New Relative Context dan SVM. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, *5*(3). https://doi.org/10.22146/jnteti.v5i3.262

Huri, M. D., Nazri, M. Z. A., Bakar, A. A., & Kurniawan, T. (2011). Reka Bentuk DNA Mengguna Sistem Imun Buatan Objektif Tunggal. *Teknologi Maklumat & Multimedia 11*, *41–48*, 41–48.

Robandi, I. (2019). Artificial Intelligence Mengupas Rekayasa Kecerdasan Buatan. In); I). Andi Off (Ed.), *Kika (ed* (pp. 183–213). ; I). Andi Offset.

Rosnelly, R. (2018). Pengenalan Pola Angka Tulisan Tangan Pada Cek Menggunakan Neocognitron. *CSRID (Computer Science Research and Its Development Journal)*, *10*(1), 23. https://doi.org/10.22303/csrid.10.1.2018.58-67

Song, M., & Sun, Z. (2014). An immune clonal selection algorithm for synthetic signature generation. *Mathematical Problems in Engineering*, *2014*. https://doi.org/10.1155/2014/324645

Yang, H., Li, T., Hu, X., Wang, F., & Zou, Y. (2014). A survey of artificial immune system based intrusion detection. *The Scientific World Journal*, *2014*. https://doi.org/10.1155/2014/156790

Zaitun, Warsito, & Pauzi, G. A. (2015). Sistem Identifikasi dan Pengenalan Pola Citra Tanda-Tangan Menggunakan Sistem Jaringan Saraf Tiruan (Artificial Neural Networks) Dengan Metode Backpropagation. *JURNAL Teori Dan Aplikasi Fisika FMIPA Universitas Lampung*, *03*(02), 93–101.

Zhang, R., & Xiao, X. (2018). A Clone Selection Based Real-Valued Negative Selection Algorithm. *Complexity*, *2018*. https://doi.org/10.1155/2018/2520940