Hemoglobin Measurement Monitoring Tool Using K-Nearest Neighbor (KNN) Algorithm Based On Internet Of Things (IOT)

Faris Hadianto[1,*] Rendy Munadi[1] Sussi[1] Arief Pambudi[1]

[1]*Bachelor in Telecommunication Engineering, School of Electrical Engineering, Telkom University, Indonesia*
[*]*Corresponding author. Email: farishdnt@student.telkomuniversity.ac.id*

ABSTRACT
Hemoglobin is a protein in red blood cells that functions as an oxygen-binding agent and also circulates to all organs and tissues of the human body. Changes in hemoglobin levels, both low and high, can indicate health problems. Checking hemoglobin levels in the human body is usually done medically by taking blood samples that are injuring the patient or often called invasive. Therefore, a non-invasive method is made, which is a monitoring tool for measuring hemoglobin using the K-Nearest Neighbor algorithm based on the Internet of Things. Internet of Things (IoT) is a promising paradigm to provide massive wireless connections in future communications. One main feature of IoT is short packet communication (SPC), which adopts finite block-length codewords for data transmissions. Different from the long packet transmission in conventional wireless networks, from the test results, it can be seen that the monitoring tool is connected to Cloud Thingspeak, the prediction process of the algorithm in python programming to the android application that runs well. The input value of Oxygen Saturation (SpO2) is carried out by the KNN algorithm processing where this algorithm is one of the Machine Learning algorithms used to predict an output value so that an average accuracy of 93.481133% at k = 2 is obtained. Quality of Service (QoS) is quantitatively defined in terms of guarantees or bounds on certain network performance parameters. The most common performance parameters for the Internet of Things (IoT) are Delay End-to-End and Throughput. Delay End-to-End has an average value in the process of sending from monitoring tools to the Android application of 6.8 s. And the average value of throughput on NodeMCU to Cloud ThingSpeak communication is 9.98 Kbps, Cloud - Python - Cloud communication is 13.76 Kbps and communication on Cloud ThingSpeak to Android Application is 11.06 Kbps.

Keywords: *Hemoglobin, SpO2, Internet of Things, K-Nearest Neighbor Algorithm, Quality of Service*

1. INTRODUCTION

Hemoglobin can be found in red blood cells that carry oxygen (O2) from the lungs throughout the body and carry blood cells containing carbon dioxide (CO2) back to the lungs. Hemoglobin levels can be said to be normal if the hemoglobin concentration in the blood is at 13.5 to 17g / dl for men and 12 to 15 g / dl for women. Disease caused by a deficiency or excess blood hemoglobin can cause various diseases such as anemia due to lack of hemoglobin and polycythemia due to excess hemoglobin. Based on data from the World Health Organization (WHO), half of the people with anemia are caused by iron deficiency. Laboratory researchers carry out a variety of physical, chemical, psychological and genetic methods to contribute greatly to clinical researchers aimed at studying various types of disorders of hemoglobin [1]. This is a very efficient way to be able to take a patient's blood sample to make it easier in the treatment process while knowing the level of hemoglobin itself.

This study is the second study from [2]. By developing this tool it can help activities in the world of health more efficiently because the process of taking Hb no longer uses blood but uses the oxygen saturation equation (Spo2) and Hb processing can be seen in that minute with a more modern output based on an android application. In the process of designing this monitoring tool, an invasive search for SpO2 and Hb training data in the Telkom University area was also carried out using the *Quick Check* tool. From this process, 56 training data for women and men were obtained.

To facilitate the measurement of hemoglobin levels repeatedly and quickly without injuring the patient's body parts can be measured non-invasively. According to research [3], techniques for measuring total levels of hemoglobin which are invasive, cannot provide fast results. This is because the technique is less effective to use when measuring hemoglobin levels. And the study conducted a measurement technique in the form of the K-Nearest Neighbor (KNN) algorithm which refers to research [4] to find the results of the approach value of these measurements by conducting training data that will be processed and entered on the android application side. The author got an idea to make a Hemoglobin measurement tool using a NodeMCU microcontroller, the MAX30100 sensor, the use of the KNN algorithm in python programming, and network side measurements namely Quality of Service (QoS).

2. BASIC THEORY

2.1 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a supervised learning algorithm that is very popular in both Data Mining and Statistics because of its simple implementation and performance for carrying out significant classifications. In Research [4], this algorithm is divided into 2 types of supervised learning, namely regression and classification which will predict a particular object based on the training data approach of several samples. Following in Table 1 are some differences from regression and classification.

Table 1. The difference between regression and classification

	Regression	**Classification**
Predict	Number	Class
Output Type	Continuous	Discrete
Model	Best Fit Line	Decision Boundary
Evaluation Metrics	*Mean Square Error* (MSE)	Accuracy

The system in this KNN algorithm will test to get a good performance. Tests will be carried out on invalid Hb data with previous data searches. The parameter that will be changed in testing is by doing a regression process of the value of k = 1 to k = 10 which is the most important part in finding the accuracy value of the tool.

In research [5], to find out the closest sample value is needed a distance calculation to find similarities and differences between point data. The purpose of this distance calculation is to find the appropriate distance or similar. And here are some distance calculations used in the algorithm This KNN:

2.1.1 Euclidean Distance

K-Nearest Neighbor can be calculated using Euclidean Distance. This is a calculation of the distance between two deep points. This is the Euclidian distance calculation in Equation 2.

$$D(x,y) = \sqrt{\sum_{k=1}^{m}\left(x_{ik} - x_{jk}\right)^2} \quad (2)$$

2.1.2 Chebyshev Distance

This distance calculation is also called the maximum value distance. This calculation can be calculated as the magnitude of the absolute value of the difference between the coordinates of a pair of points from data. And here is the Chebyshev distance calculation in Equation 3.

$$D(x,y) = \max_k \left|x_{ik} - x_{jk}\right| \quad (3)$$

2.1.3 Manhattan Distance

This calculation is useful for calculating the difference between the coordinates of a pair of data points. Following is the calculation of Manhattan's distance in Equation 4.

$$D(x,y) = \left|x_{ik} - x_{jk}\right| \quad (4)$$

2.2 Hemoglobin

Hemoglobin is a protein molecule in the blood that can bind oxygen. One very important indicator in the supply of oxygen in the body is oxygen saturation (SpO2). Oxygen saturation (SpO2) is a percentage of oxygen levels distributed throughout the body that has been carried by HbO2. Following is the relationship of SpO2 and Hb in Equation 5 [6].

$$SpO2 = \frac{HbO2}{HbO2+Hb} \quad (5)$$

2.3 Pulse Oximeter (MAX30100)

The MAX30100 sensor is a pulse oximeter in the form of a heart rate monitor and an Oxygen Saturation (SpO2) sensor module. In Research [7], this sensor has a red LED with a wavelength of 660nm and an infrared LED with a wavelength of 880nm. The workings of these sensors in the form of detection of light in the form of each on the side of the probe that is useful for spreading light to the light detector through body tissues. Following in Figure 1 which is the workings of these sensors:

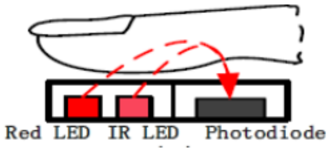

Figure 1 Detection scheme of the sensor pulse oximeter.

The two LEDs will be accompanied by photodiode light which will be absorbed and will continue the process of converting analog to digital filters. In this case, the sensor will be carried out on a non-invasive basis. This sensor will be connected to a microcontroller for processing sensor data. Following Figure 2 is a comparison of red LEDs and infrared LEDs with Hb and HbO2:

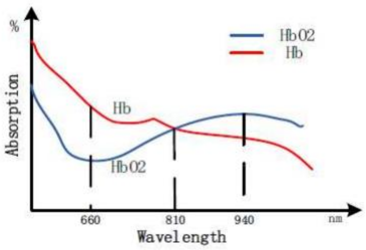

Figure 2 Graphic relationship between infrared and red LED with Hb and HbO2.

This sensor uses I2C communication for communication on two ICs using Serial Data (SDA) and Serial Clock (SCL) pins. The SCL pin is used to regulate and control the sending of sensor data to the microcontroller which is sent by the sensor via SDA [8].

2.4 Cloud ThingSpeak

ThingSpeak is an open-source IoT platform that allows being able to collect, visualize, analyze data directly, and be able to react according to that platform. This platform can help to build an IoT system without the need to set up additional servers. Data collection is done using the MQTT protocol and can use the HTTP protocol. The main component of ThingSpeak is a channel which can store data transmission from various devices [9].

Hyper Text Transfer Protocol (HTTP) is a communication protocol that can connect clients, servers and IoT devices. In HTTP the sending pattern has a Request and Respond scheme [10].

2.5 Quality of Service

Quality of Service (QoS) is a method for measuring and defining how good the nature of a service is. A set of work attributes that have been associated with a service can be measured with QoS. Therefore, Qos can also be interpreted as a network resource management technique. In QoS, there are several measurement parameters such as throughput and delay [11].

2.6 Internet of Things

Internet of Things (IoT) is a concept that aims to expand the benefits of continuously connected internet connectivity [12]. Basically, the Internet of Things (IoT) refers to objects that can be uniquely identified as virtual representatives in internet-based structures.

How the Internet of Things (IoT) Works is the interaction between machines that are automatically connected without user intervention and at any distance. In order to achieve the workings of the Internet of Things (IoT), the internet is needed as a liaison between the two machine interactions. The benefit from the concept of the Internet of Things (IoT) itself is that the work to be done can be faster, easier and more efficient.

3. METHODOLOGY

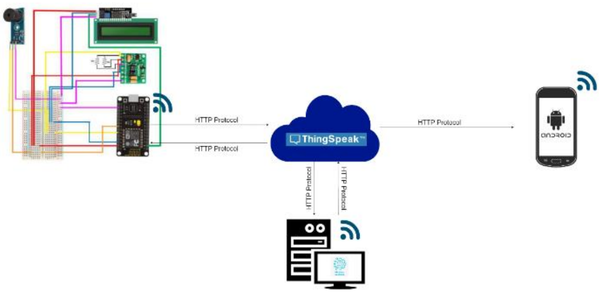

Figure 3 Design system of Hb monitoring tool

Measurement of hemoglobin levels using a non-invasive system where the process of checking hemoglobin levels does not need to use blood samples. Figure 3 is a 4 element system design that will be used namely the microcontroller, Cloud ThingSpeak, server, and Android Applications. The protocol used in the communication process of the four elements is the HTTP protocol. The MAX30100 sensor will send the SpO2 value from the transfer process of the Infrared LED Reflective Light and the red LED which will be absorbed by the photodiode light. The sensor value sent to NodeMCU is already in the form of SpO2. When the SpO2 value is greater than 0, there will be a reading process on the LCD screen and a Buzzer that will sound when the SpO2 value is successfully sent to Cloud ThingSpeak. Sending data from the microcontroller will fill 2 fields which are monitoring graphs on Cloud ThingSpeak. The first field is the SpO2 value field and the second field is the Hb value field with a value of 0. Then the server will process using python programming and update the value in Cloud ThingSpeak if the value of the first field is greater than 2 and the second field is 0. In the Cloud ThingSpeak, the value to be read is the value that was last updated. After the condition is accepted by the server, the process of predicting values based on the regression of the k value obtained during the training of the K-Nearest Neighbor (KNN) algorithm will give the results of the Hb value which will be updated in the second field and the value of 0 for the first field. On the Android application side, it will read the value of the second field and define the status of the body condition when the user checks the Hb level.

Testing in the process of monitoring tools will be carried out by analyzing a hemoglobin level value that comes out of a tool that has been designed in a non-invasive manner by comparing the hemoglobin measurement tool which is more accurate and invasive, namely the Quik-Check tool, so that the percentage error rate of the tool can be found. non-invasive before using an algorithmic approach. And the following categories of hemoglobin levels for adults are in Table 2.

Table 2. Classification of Hemoglobin (Hb) Levels

Gender	Category	Gender	Category
	The Lack of Hb		**The Lack of Hb**
Adult Male	<13,5 g/dl	Adult Male	<13,5 g/dl
Adult Female	<12 g/dl	Adult Female	<12 g/dl

Delay is the time it takes for data to travel the distance from origin to destination. In this case, the delay is influenced by the time that lasts a long process, distance and physical media. Table 3 shows the Delay category according to the amount of delay.

Table 3. Category of Delay

Category	Delay (ms)
Excellent	< 150 ms
Good	150 ms – 300 ms
Satisfied	300 ms – 450 ms
Bad	> 450 ms

Source: TIPHON

Throughput is an effective and measurable data transfer in bits per second (bps). Throughput is calculated as the number of successful packet arrivals observed at the destination when a certain time interval is divided by the duration of that time. Throughput categories are shown in Table 4.

Table 4. Category of Throughput

Category	Throughput (bps)
Excellent	>2,1 Mbps
Good	1200 Kbps – 2,1 Mbps
Satisfied	700 – 1200 Kbps
Bad	338 – 700 Kbps

Source: TIPHON

4. RESULTS & ANALYSIS

Based on the results of experiments that have been carried out on the parameters k = 1 to k = 10 according to Table 5. From these data, it can be seen that the highest average level of accuracy is at k = 2 of 93.481133%.

Table 5. The overall average value of each parameter k

KNN		Accuracy(%)
k=	1	92,7025611
	2	93,4871133
	3	93,3802562
	4	92,6299471
	5	92,2012265
	6	92,8351192
	7	93,0177505
	8	92,9548548
	9	93,0185318
	10	93,0185318

From Table 5, we can see that k = 9 and k = 10 have the same accuracy value. This means that the greater the parameter k, the accuracy obtained will be of fixed value because of the limited training data. As more training data is obtained, the accuracy of each parameter k, whether k = 1 or more will be unstable. So the parameter k will be used in the Hb monitoring tool.

Table 6. Hb accuracy at k=2

SpO2	Hb Invansive	Hb Non-Invansive	Difference	Accuracy(%)
98	15.9	14,3	1,6	89,9371069%
93	13.9	13,95	0,05	99,6402878%
96	13.8	13,55	0,25	98,1884058%
97	11.9	13,6	1,7	85,7142857%
99	12.7	13,8	1,1	91,3385827%
97	13.2	13,6	0,4	96,9696967%
92	12.5	14,1	1,6	87,2%
95	13.6	13,5	0,1	99,2647059%
99	14.7	13,8	0,9	93,877551%
94	15.2	15,2	0	100%
97	14.2	13,6	0,6	95,7746479%
94	16.5	15,2	1,3	92,1212121%
97	14.5	13,6	0,9	93,7931034%
97	16	13,6	2,4	85%

Table 6 is about the Hb accuracy value that has been processed from the algorithm that has been made and the non-invasive Hb value is obtained from the spo2 input data on the MAX30100 sensor which will later be processed to predict the Hb value.

Delay measurements in the monitoring tool system are calculated by the length of time the data takes from origin to destination. The delay will be calculated from the process of SpO2 value sent by NodeMCU until the Hb value is sent to the Hb application, a description of the delay measurement sequence scheme calculated from the NodeMCU – Cloud – Server – Cloud – Android Application. This process was carried out as many as 7 sessions, where each session conducted an experiment 10 times.

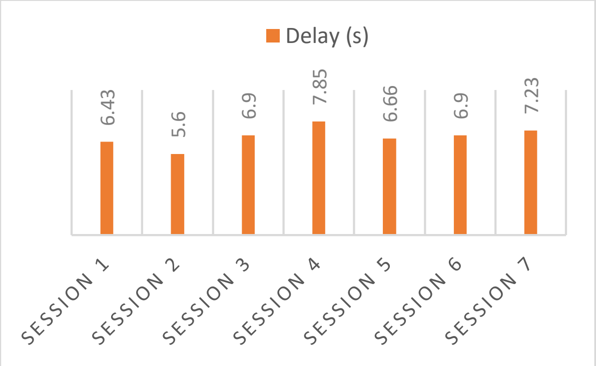

Figure 4 Delay End-to-End monitoring system diagram for 70 trials

Based on Figure 4 there is a delay value for each session, where the largest average delay value is in the 4th session which is 7.85 s or 7850 ms. Thus also the average value of the fastest delay in the monitoring tool system was found in the second session of 5.6 s or 5600 ms. This is influenced by the process of predicting the Hb value which takes time when the SpO2 value enters the cloud, the value must be sent at once it needs to be processed again by python programming using the KNN Algorithm before sending it to the Android Application. This delay is also affected by the HTTP protocol which will process requests and replies that require time too. However, with a delivery delay of 6.8 s, there is no effect on the Hb monitoring tool that has been designed.

Throughput measurements are performed on 3 communication system models, namely NodeMCU–Cloud communication, Cloud–Server–Cloud communication, and Cloud–Application communication. Each communication model has carried out as many as 70 experiments divided into 7 sessions.

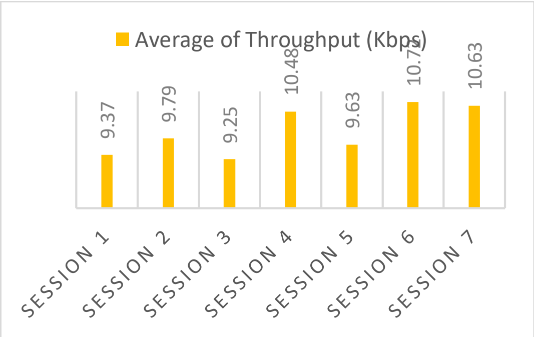

Figure 5 Diagram of average throughput on NodeMCU - Cloud communication

Figure 5 is the average throughput value from the NodeMCU ESP8266 communication to the process to Cloud ThingSpeak for sending SpO2 values. Throughput testing is carried out as many as 7 sessions where each session is carried out as many as 10 attempts.

Wireshark application is used to check the value of throughput by filtering IP NodeMCU ESP8266 which is 192.168.137.170 with the destination that is to Cloud ThingSpeak which has an IP of 52.2.53.80. In this test, the highest average throughput is found in session 6, which is 10.72 Kbps and the smallest average value is in session 3, which is 9.25 Kbps.

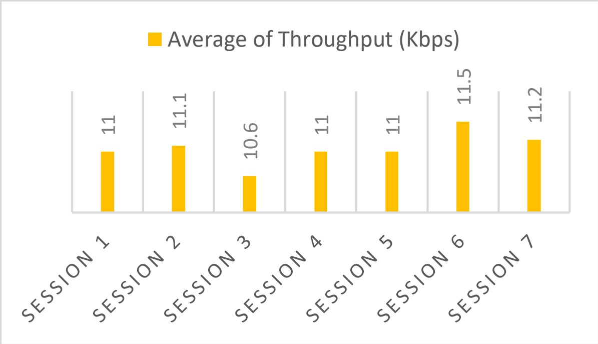

Figure 6 Diagram of average throughput on Cloud - Python - Cloud communication

Figure 6 is the average throughput value from the Cloud ThingSpeak communication to the processing to the server and the process of sending Back to Cloud ThingSpeak. There are 2 types of data sending processes, namely the SpO2 value and the Hb value. This test shows that the highest average throughput is in session 6, which is 15.2 Kbps and the smallest average value is in session 3, which is 12.9 Kbps.

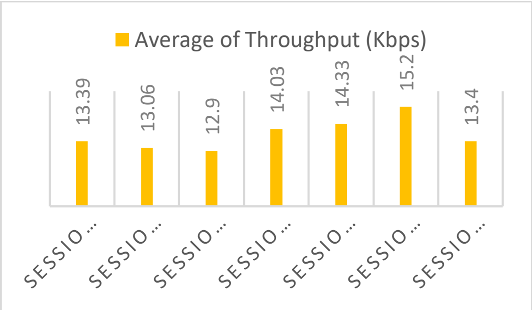

Figure 7 Diagram of average throughput on Cloud communication - Android application

Figure 7 is the average throughput value from the Cloud ThingSpeak communication to the Android Application. this process will send the Hb value to the Android Application. This test shows that the highest average throughput is in session 6, which is 11.5 Kbps and the smallest average value is in session 3, which is 10.6 Kbps.

Based on all this average value of throughput is influenced by the system that sends 2 data at one time sending to Cloud ThingSpeak as a Hb monitoring tool scenario. In Figure 5, the data sent from NodeMCU to Cloud ThingSpeak is the SpO2 value in field 1 and the value 0 in field 2. In Figure 6, which is a Cloud – Server – Cloud communication data that is sent to Python programming as well as 2 data, namely the SpO2 value in the field 1 and value 0 in field 2 and sent back to Cloud ThingSpeak with field 1 value 0 and field 2 are the Hb values that have been processed so that in this process there is the largest average value of throughput compared to the two other system communications, and for Cloud communication ThingSpeak until the Android Application according to Figure 7, data sent will any change in field 2 or the processed Hb value.

5. CONCLUSIONS

Based on the results of testing that has been done, the authors can draw conclusions as follows are The process of designing a non-invasive hemoglobin monitoring tool using the IoT-based KNN algorithm runs well; The input value on the SpO2 with the KNN algorithm processing produces the highest average accuracy of 93,481133% at k = 2; The greater the specified k value, the percentage of algorithm accuracy is more stable due to having very limited training data collected; The average End-to-End Delay in the process of sending the SpO2 value up to the Hb value to the android application is 6,8 s or 6800 ms; and The average throughput on NodeMCU communications to Cloud Thingspeak is 9,98 Kbps, Communication Cloud – Server – Cloud is 13,76 Kbps and communication on Cloud Thingspeak to Android Applications is 11,06 Kbps.

REFERENCES

[1] K. S. Pavithra, X. Anitha Mary, K. Rajasekaran, and R. Jegan, “Low Cost Non-Invasive Medical Device for Measuring Hemoglobin,” Proc. IEEE Int. Conf. Innov. Electr. Electron. Instrum. Media Technol. ICIEEIMT 2017, vol. 2017-Janua, pp. 197–200, 2017.

[2] A.Pambudi, “Alat Monitoring Hemoglobin menggunakan Algoritma jaringan saraf tiruan propagasi kembali berbasiskan Internet of things,” Bandung Univ. Telkom, 2020.

[3] S. Chugh and J. Kaur, “Non-invasive hemoglobin monitoring device,” 2015 Int. Conf. Control. Commun. Comput. India, ICCC 2015, no. November, pp. 380–383, 2016.

[4] S. Imandoust and M. Bolandraftar, “Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background,” Int. J. Eng. Res. Appl., vol. 3, no. 5, pp. 605–610, 2013.

[5] P. Mulak and N. Talhar, “Analysis of Distance Measures Using K-Nearest Neighbor Algorithm on KDD Dataset,” Int. J. Sci. Res., vol. 4, no. 7, pp. 2319–7064, 2015.

[6] M. Rezza et al., “Perancangan dan Implementasi Alat Ukur Kadar Hemoglobin Darah secara Non-Invasive berbasis Arduino,” Bandung Univ. Telkom, 2018.

[7] J. Wan, Y. Zou, Y. Li, and J. Wang, “Reflective type blood oxygen saturation detection system based on MAX30100,” 2017 Int. Conf. Secur. Pattern Anal. Cybern. SPAC 2017, vol. 2018-Janua, no. 4, pp. 615–619, 2018.

[8] Maximintegrated, “Max30100,” Lect. Notes Energy, vol. 38, pp. 11–37, 2014.

[9] D. Nettikadan and S. Raj, “Smart Community Monitoring System using Thingspeak IoT Plaform,” Artic. Int. J. Appl. Eng. Res., vol. 13, no. September, pp. 13402–13408, 2018.

[10] Bharati Wukkadada; Kirti Wankhede; Ramith Nambiar; Amala Nair, “Comparison with HTTP and MQTT In Internet of Things (IoT) - IEEE Conference Publication,” 2018 Int. Conf. Inven. Res. Comput. Appl. (ICIRCA 2018), no. Icirca, pp. 249–253, 2018.

[11] R. Wulandari, “Analisis Qos (Quality of Service) Pada Jaringan Internet (Studi Kasus : Upt Loka Uji Teknik Penambangan Jampang Kulon – Lipi),” J. Tek. Inform. dan Sist. Inf., vol. 2, no. 2, pp. 162–172, 2016.

[12] S. M. P. P. S. A. P. Keyur K Patel, “Internet of Things-IOT Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges,” Ijesc, vol. 6, no. 5, p. 10, 2016.