



Cry Recognition for Infant Incubator Monitoring System Based on Internet of Things using Machine Learning

Erwin Sutanto^{1*}Fahmi Fahmi²Wervyan Shalannanda³Arga Aridarma⁴

¹Biomedical Engineering, Fakultas Sains dan Teknologi, Universitas Airlangga,
Kampus C Mulyorejo, Surabaya 60115, Indonesia

²Department of Electrical Engineering, Universitas Sumatera Utara, Medan, Indonesia

³School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia

⁴PT. Tesena Inovindo, Jakarta, Indonesia

* Corresponding author's Email: erwin_sutanto@fst.unair.ac.id

Abstract: With the current technology trend of IoT and Smart Device, there is a possibility for the improvement of our infant incubator in responding to the real baby's condition. This work is trying to see that possibility. First is by analyzing of open baby voice database. From there, a procedure to find out baby cry classification will be explained. The approach was starting with an analysis of sound's power from that WAV files before going further into the 2D pattern, which will have features for the machine learning. From this work, around 85% accuracy could be achieved. Then together with sensors, it would be useful for infant incubator's innovation by utilizing this proposed configuration.

Keywords: Infant incubator, Internet of things (IoT), Machine learning, Convolutional neural network (CNN), Tensorflow.

1. Introduction

Thermoregulation should be a significant problem in infant incubator. The target of body temperature for infants on average is around 37 C [1]. In general, this is for newborn babies, especially premature babies who have weak internal thermal regulation control to maintain their body temperature. It could be due to the insufficient energy reserves of the baby itself, which might not be enough to produce body heat. For this, supports are needed from external heat production, which could fit into the baby's condition. A common way to help babies regulate their body temperature back to normal is using the incubator. It keeps the baby warm and provides the controlled heat so the baby could quickly adapt to the conditions instead of the outside environment [2]. As most concern for infant incubator has been into the temperature and humidity [3]. These factors were answering through the medical perspective for premature infants. It is preventing the risk of

hypothermia by following the International Liaison Committee on Resuscitation (ILCOR) guidelines [4]. However, it should be noted that since the early development of infant incubator [5], the electric power source has been used for heating source to generate steam [5, 6].

That is why it may need a large number in every health centre. However, this is sometimes not in linear with the number of nurses in hospitals, especially in the remote area. Thus, a better incubator might be needed to help the nurse. Even more, technology has been more sophisticated than before; the electricity usage is getting much more. The next matter should be about how to use it safely. At least, the controller would be able to detect the need for immediate actions in kind of electric shocks in medical devices [7]. Hence, it should also be made sure that it could prevent any accidents. That means besides its operational function, and it should also meet the safety standard. Its possible danger can also be listed from the infant incubator analyzer [8]. For example, it could happen in a case of the medical

device misuse of the tool or any human mistakes caused by operators.

Nowadays, more and more Internet of Things (IoT) devices are coming up with recent available technology. This particular technology really improves on how communication between devices happened [9]. This improvement is not only making it more accessible; it also makes it possible for further control improvement with current technology. One example is for emotion recognition in speech using hybrid deep network scheme [10]. Besides that, there are many possible applications for this new kind of technology. It will be possible also to apply it for most of the basic measurement using common sensors. For example, if we are looking back at infant incubators, it could use a temperature sensor. It could be used either for the incubator and also the body temperature of the baby.

The purpose of doing this design is to create a better monitoring device for the incubator, which could help the nurse. The common way to do this is by interpreting the baby’s cry. The benefit from this is to reduce the frustration of unknowing what to do in a real problem. In many cases, a nurse could be guided by machine learning in order to understand the babies.

Common infant incubators would be used as a regulator of temperature and humidity only. This could be improved to have more functions, such as analyzing air condition inside the incubator in order to find the best condition for the baby’s health. It could also use a webcam to monitor Baby activity [11]. In order to help the nurse, a crying decoder should be helpful enough to calm and nurse the baby [12]. An IoT system would be used here to provide the communication facility. It will help in collecting the data which would be stored at a server; it might be further processed automatically by using machine learning in that server. However, in this work, we only focus on cleaned baby voice out of the background [13]. This will help on classifying the cry with all possible parameters extracted using python.

2. Methods

By seeing the importance of the condition of the baby concerned then, there are three things which we should like to discuss. They are the infant incubator, the baby voice itself, and the possible monitoring system design. The monitoring system design will include possible sensors to improve infant incubator.

2.1 Infant incubator

An incubator is a device used to treat and protect newborn babies with premature or unusually small in



Figure. 1 Tesena infant incubator

size. It offers a controlled environment, especially temperature. PT Tesena Inovindo produced the incubator used in this experiment. It offers closed-loop temperature control, with selection between measured skin temperature or chamber temperature as a feedback value for the system.

The chambers temperature is maintained by means of circulating air through a heated air duct underneath the baby chamber. Depending on the selection, the resulting temperature is continuously monitored, and then the heating is adjusted according to the users setting value. This incubator also features a means to add humidity inside the chamber by setting the window size of a water container manually. Water evaporates when heated air passed through its surface and carries humidity to the chamber.

2.2 Baby voice

The common method to interpret the baby’s voice is using Dunstan Baby Language (DBL). The language categorized the voice into five different words. The categories are shown in Table 1. Analyzing the baby’s voice is easier using available data, especially a labelled dataset. Five Baby’s word from the DBL actually could be used with the Open database from “Donate a cry corpus” [14]. From the number of WAV files, it could be seen that the distribution of data over each category was not even. There was more data in the category of hungry than other categories. It is more than 83% of the overall data.

Table 1. Baby voice category

Word	Meaning	Folder	∑ Files
Neh	I am hungry	Hungry	382
Owh	I am sleepy	Tired	24
Heh	I feel uncomfortable	Discomfort	27
Eairh	There is air in my stomach	Belly Pain	16
Eh	I want to burp	Burping	8

Table 2. Baby’s profile

Reason of cry	Gender		Age				
	Boy	Girl	0-4w	4-8w	2-6m	7m-2y	> 2 y
Hungry	232	150	133	42	130	68	9
Discomfort	15	12	6	7	11	2	1
Tired	11	13	6	3	13	2	0
Belly pain	15	1	3	1	4	8	0
Burping	3	5	1	2	4	1	0

* w: weeks old, m: months old, y: years old

The database was coming from the recorded voice of a baby using mobile phones. They were coming from many people who would like to contribute. As the voice itself was not from one baby, the recorded files have indicated with their profiles, such as gender, age group, and reason for crying. Composition details of WAV files over their profiles could be seen in Table 2.

From Table 2, we could see furthermore that the distribution of data based on age profiles was not even also. There were more data in 0-4 weeks old than other groups of age in the category of hungry. It is about 133 over 382 of hungry data. It is about 34.8%. Luckily, most numerous data is also the most important group of age for infant incubator. Usually, babies who would be using the incubator were also in that group of age. Babies must grow to their ideal weight within weeks and return to his mother. The total of data must be enough to accommodate the training, validation, and testing as in Eq. (1).

$$\sum data = \sum train + \sum validate + \sum test \quad (1)$$

For consideration, if we just use 0-4 week old group as the limit from the incubator, the other non-hungry data within this age group was only 16 data. It was only about 11%, and it still had to be also used for other purposes such as training, validation, and testing. Thus, our baby’s voice recognition should use this dataset carefully as the number was not too many. We could also use other age groups with a number of considerations.

2.3 Intelligent module based on internet of things

Fig. 2 is also showing the possible sensors for the chamber. The LM32 module could be used for the temperature while the DHT11 module might function as humidity measurement. Both sensors would become feedback for the heater and also inform the operator to increasing or decreasing the window size of a water container as it is provided manually. Extra sensors, both MQ-3 and MQ-135, would be indicating for gas composition inside the chamber

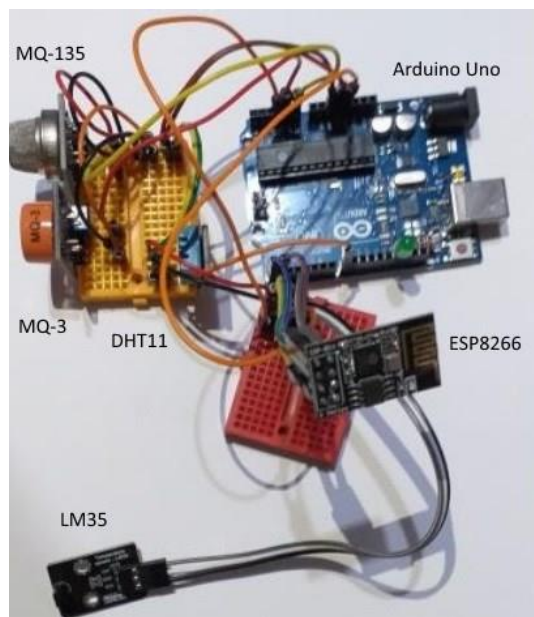


Figure. 2 Hardware module

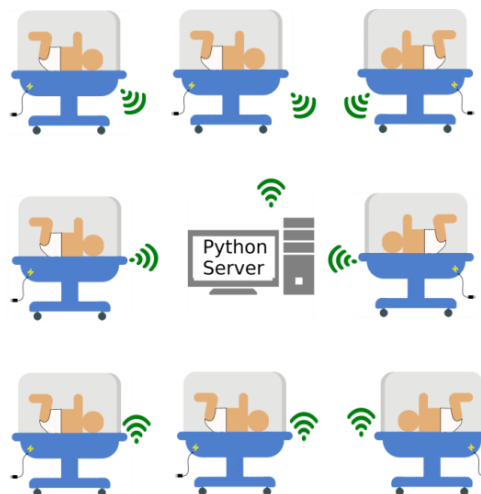


Figure 3. Overall analysis system

such as NH₃, alcohol, benzene, smoke, CO₂, et cetera. It could give a hint if the incubator were open for air circulation. The data will be sent to the server via the IoT module so that it can be processed and displayed the desired data. Those sensors would help to keep the best environment for the baby.

Besides some possible sensors, there is also Arduino Uno as a common microcontroller module and ESP8266 for its communication module, as shown in Fig. 2. All possible data, including the baby’s voice, will be collected into a server for further analysis to get comprehensive data of the baby. The baby’s crying would help to indicate data from the incubator’s chamber and also the baby condition itself. The monitoring system will be then possible to justify the baby’s real environment from chamber temperature and its effects on the baby from the skin temperature and the baby voice. As discussed, the

communication could be utilizing the IoT technology. From those various data, the best condition for the baby could be maintained. That will include a process of sending data to the IoT Server and returning response values for the actuators of the incubators.

After the implementation of that IoT platform, there is a possibility for calculation and data analysis in server. It will be processing using the data using an algorithm periodically and be concluded based on collected data of baby and incubator data. It could use a machine learning technique in order to find out the data pattern. A complete system of this is shown in Fig. 3.

Fig. 3 shows where the incubator is monitored by a python-based web server. The server would use the available python libraries to extract the feature from the baby voice. After all, the sensor's data and actuators value would be just recorded in a database. Before being sent to the web interface as a report to users for further manual actions, this system will perform machine learning-based analysis, including baby voices, and having the historical data of each incubator. How to record the data would be obvious. However, how to understand the baby voice would need a kind of voice algorithm. The common algorithm for interpreting voices is using Mel-Frequency Cepstrum Coefficients (MFCC). The method is using the following steps to extract the voice features.

$$voice_{window}(n) = voice_{digital}(n) * w(n) \quad (2)$$

The first step is using the windowing technique. With $0 \leq n \leq N - 1$ as the index of $voice_{digital}$, $voice_{window}$ could be calculated by using its convolution with the window function $w(n)$. The digital voice is recorded voice of WAV files. It could be one from the database at Table 1.

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$

The well-known function is the hamming window at Eq. (3). With this function, we could process various length of files sequentially through a small window of time length. This is again a common method for Finite Impulse Response (FIR) Filter Design.

$$X_n = \sum_{k=0}^{N-1} x_k \cdot e^{-2\pi jkn/N} \quad (4)$$

After that, the process itself could be done in the frequency domain using a Discrete Fourier Transform (DFT) as in Eq. (4). That is with n as the index of the signal and k as index following the summation operation. This is related to Mel-

Frequency which has rules for its conversion as follows.

$$mel(f) = \begin{cases} 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) & \text{if } f \geq 1\text{kHz} \\ f & \text{if } f < 1\text{kHz} \end{cases} \quad (5)$$

With f as the signal frequency in Hz, $Mel(f)$ is the scale for new frequencies value conversion to get the feature from its original spectrum [15]. This new spectrum is based on the human hearing range, which is at 2000 - 5000 Hz frequency range.

$$X_i = \log_{10}\left(\sum_{k=0}^{N-1} |X(k)| \cdot H_i(k)\right) \quad (6)$$

Using i as the filter index with a value of $1, 2, 3, \dots, M$, and H_i is the triangle filter at index i , we could calculate X_i as the convolution of the voice signal $X(k)$ and the filter $H_i(k)$. The new frequency spectrum will be used to the MFCC features.

$$C_j = \sum X_i \cos\left(\frac{j\pi(i-1)}{2M}\right) \quad (7)$$

MFCC features c could be extracted as in Eq. (7) by having j as a coefficient index with a value of $1, 2, 3, \dots, J$, and M is the number of features. With the values, we could then construct the spectrogram as a representation object of the voice data.

$$g[m, n] = (w * f)[m, n] = \sum_j \sum_k w[j, k] f[m - j, n - k] \quad (8)$$

Convolutional Neural Network (CNN) would be an application of convolution as in Eq. (8) and the features similarly like Artificial Neural Network (ANN) with weighing factor ω . This multiplication makes the possible calculation for input 2D dataset f with a size of $m \times n$.

$$T = [T^{e_1} \quad T^{e_2} \quad T^{e_3}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \quad (9)$$

Finally, as we use python, it is possible to use developed libraries for MFCC [16] and Tensorflow [17] to accommodate the CNN calculation. Tensorflow will be useful to work with the vectors as in Eq. (9).

Parents and nurses as decision-makers can assess the condition of the baby in the incubator and produce complete information about the baby. As the data is more into unreadable by the user, it is expected to use the described method to conclude them in an

understandable form. The data could be then accessed from the web or mobile app.

3. Result and discussion

From the recording shown in Table 1, it is possible for us to get MFCC features and apply CNN to learn the characteristic of a baby’s cry with some limitations. This is due to the fact that from five baby classification, namely; hungry, tired, discomfort, belly Pain, and burping, it is only hungry, which has enough data. After the baby voice, it is not difficult to see all the sensor data via the database.

3.1 Baby voice

The first validation of open cry database [14] was using the length of the records. It made sure that all the data had enough length. Total data was 457. The plot of overall data could be seen in Fig. 4.

Fig. 4 showed that the range of time length was between 6.0-7.1 seconds. Based on the range, the window time could be 50 msec and the used time length is 6.30 seconds. This specification will be used later to generate mfcc features. From the figure, it could also be seen that the hungry’s category with green color, had dominated most of the data set. It could be seen that the other symbols were distributed within the range but none as many as green color. Based on its power value on the vertical axis, it could be noticed that overall categories have almost similar value range. The maximum value of hungry category around 12,000 W was closed by discomfort value. Meanwhile, the mean value of each category also showed similar value which is around 1,000 W with a little different for each category.

Although the data was not balanced for each category, there is still a possibility to find out the characteristic of power in the expectation that it is different between each category. In that regard, the calculation should check all the WAV files for power value on each category. From there, we could see the

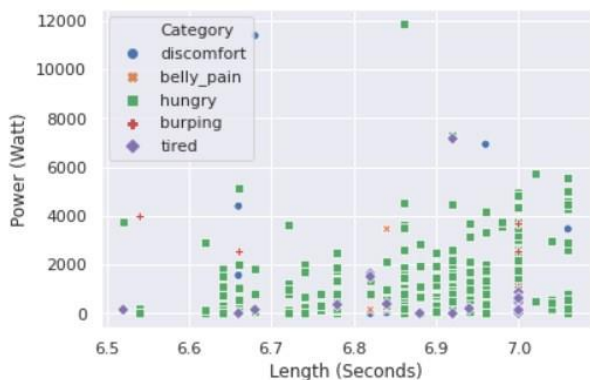


Figure 4. Distribution of power over record’s length

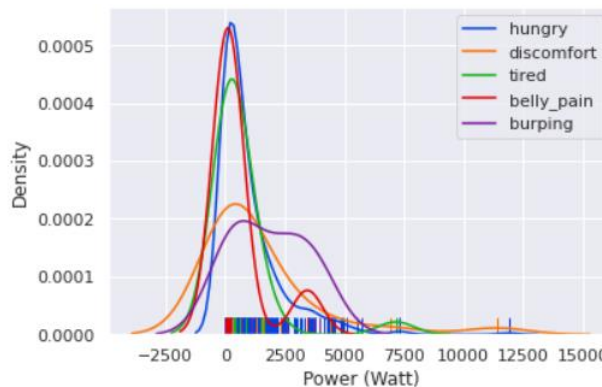


Figure 5. Comparison of power’s distribution

distribution of the power from overall data. The distribution of power on each category was shown at Fig. 5.

Fig. 5 showed the distribution. From the figure, it could be seen that the maximum probability of power value from each category was about the same. Although the hungry category has a less of power value compared with other categories, it does not mean much. Looking again into the shape of the distribution, we could see it is different compared with tired, discomfort, and burping. However, it shows that the shape of hungry has almost similar with belly pain. This may indicate that the data could not be separated only by using power value.

Looking back into the number of data, the hungry data was only left 133. It was if the concern was for infant incubator since the data would be limited to 0-4 months old only. Thus, the classification could be arranged just into two categories which are hungry and non-hungry. The total number of non-hungry was only about 75, although it was already included all data from the other four categories, such as tired, discomfort, belly pain, and burping.

Meanwhile, the number of recordings with hungry indication has more data compared to other categories. Thus, it is not possible to use it as data training for each category. It is more probable to classify the voice as hungry over non-hungry. With a total of data, as mentioned before, the hungry group has about 1.77 times more data compared with the non-hungry. Again it was with reality of non-hungry group was less than 100 data. This may affect later on learning rate with its accuracy and loss validation.

The feature of each baby voice as discussed was extracted using mfcc algorithm. By using this method, it is possible to map the feature against its windowing index to become 2D graphs. It is shown in Fig. 6. The figure showed spectrogram for each baby voice category. The color was showing the mfcc value from each recording voice using Eq. (7). From these figures, we could see that all of them were

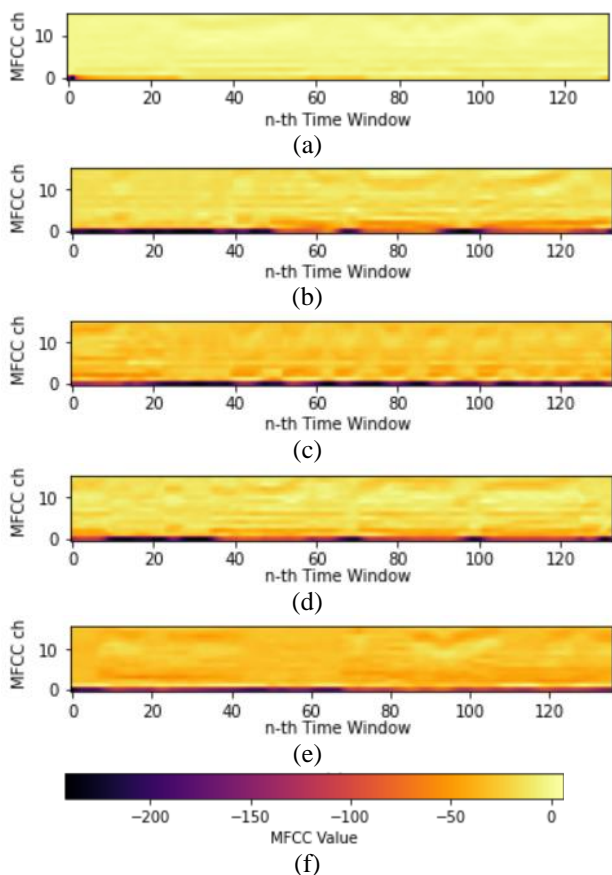


Figure. 6 Spectrogram: (a) hungry, (b) tired, (c) discomfort, (d) belly pain, (e) burping, and (f) color map

constructed in the same size height and width. The Y-axis itself was actually the scaling of MFCC frequency. In this work, we use 16 as the number of cepstrum or MFCC frequency length as discussed from Eq. (5). Then, for the time length of voice as x-axis, we limit it into 126 of Hamming Window as in Eq. (3). With the window length of 50 ms, we would have limited time length is 6.30 seconds as elaborated before from Fig. 4. From there, we could compare each other. It could be seen that hungry cry had the lightest color compared to other categories. It indicates that it has a lower intensity's value. The gradual increase of the color was shown from belly pain, tired, discomfort, and burping respectively by inspection. However, discomfort and burping were just different from the pattern.

Although from all of them, it is possible to see the difference, we still had to stick with two categories only. It is because this graph is the only representation of one WAV file. In order to make a general pattern, we have to consider all of the available WAV files. Thus, we decide to use hungry as the target of classification. Looking at those 2D graphs, there were more possibilities to find its characteristics. However, it might need different method compared last results from 1D data such as

power distribution. It will use the availability of a database and learn from there to classify the hungry voice.

3.2 Machine learning and database

To learn and to classify the voice, machine learning will be used to accommodate the number of files. Here, we could use CNN over common ANN. With this technology, it is possible to learn the files without necessary to find the characteristic manually as 1D approach of power distribution. It used a procedural method which was passing through the several phases. The first phase was making a possible model. There should be kind of several activation layers with the size of the 2D sample of our graphs which was an input shape of 16 x 126 per 1 file.

After that, there were the training phase and testing phase before going to the validation phase. To go inside those last phases, we should manage our limited number of files. Here, we just use standard proportion which is 10% for validation and 10% for testing. It will leave us only 80% for the training. This will divide our available 133 of hungry files and 75 of non-hungry files into three random groups of data. They will become 168 of training files, 20 of validation files, and 20 of testing files. All of them were randomize a group of hungry and non-hungry data. After the training, we could get the result of testing and validation, as shown in Fig. 8.

Fig. 8 shows the result of our tested model. It could get a training score of 84.52% for accuracy and 39.43% for loss. This could be improved by changing the number of layer and activation function on each of the layer in the machine learning model. However, it will still be affected by the number of data. That limit was visible from 20 of validation files as the learning stuck at around 75% for the loss and 50% for the loss. It could be seen that after epoch reached 20, the learning was going for overfitting on both accuracy and loss. This once again affected our 20 testing files which classified as in Table 3. It showed the confusion matrix from the testing file, which was in total 20. It could give out the performance of our learning model.

By having Hungry group which was detected with Positive value of 8 as True Positive (TP), then Non-Hungry one that was detected with Negative with a value of 3 as True Negative (TN), we could get its accuracy, precision and sensitivity of our CNN

Table 3. Confusion matrix

Score	Hungry	Non-Hungry
Positive	8	6
Negative	3	3

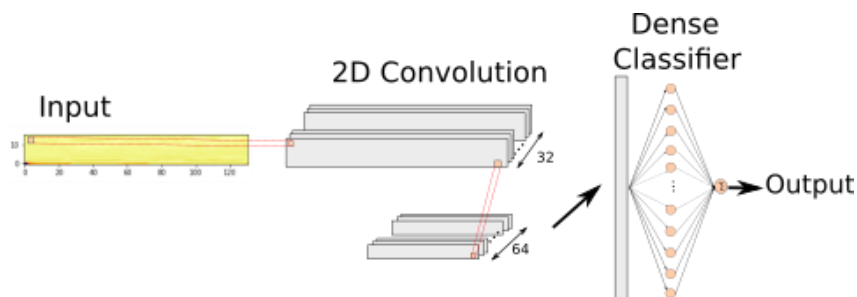


Figure. 7 CNN structure

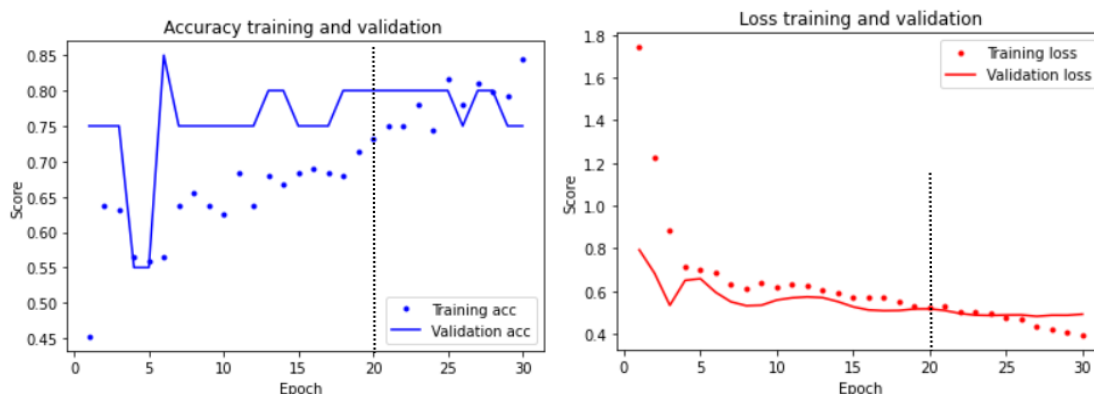


Figure. 8 Training and validation

	id	humidity	temperature	mq135reading	mq3reading
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	301	62	30.4	85	102
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	302	62	30.2	88	107
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	303	63	30.5	88	108
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	304	63	30.2	89	109
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	305	64	30.6	87	106
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	306	64	30.2	87	107
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	307	64	30.3	88	108
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	308	64	30.2	91	110
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	309	64	30.5	91	110
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	310	64	30.2	91	109
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	311	64	30.5	90	109
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	312	64	30.2	89	108
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	313	64	30.4	90	110
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	314	64	30.3	88	108
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	315	64	30.6	89	109
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	316	64	30.2	88	107
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	317	64	30.3	89	109

Figure. 9 Database view

model following the accuracy which was stated before. It was by using the following equations.

$$Accuracy = \frac{TP+TN}{P+N} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{12}$$

The other two groups were known as False Positive (FP) and False Negative (FN), which got 6 and 3, respectively. Then, we also had P as a total of Positive group and N as a total of Negative group,

which were 14 and 6 in order. The accuracy of the structure was only 55%. Meanwhile, for Precision and Sensitivity, the values were 57% and 73% in sequence.

The percentage score could be changed by repeating the process as the weighing value inside CNN layers was updated automatically. However, the overfitting trend would always be visible due to this limited number of data sets. The limited number of the non-hungry group was helping us to see the possibility of the real condition. In which the early infant below four months will mostly cry because of hungry. This is understandable because babies at this age need to double their weight within months [18]. In respect of “Donate a cry corpus” [14], our CNN model had shown better result compared with clustering method using statistical feature extraction and gaussian mixture models [12] which is having an accuracy of 81.27%. Again with the same source, the other test with the machine learning method was implemented for Automated Baby Monitoring which could reach 96% precision. However, it is only for crying/non-crying detection [11]. Thus, this approach using Tensorflow which could reach 85% from the training was promising because it could suggest the hungry condition over other categories. It was better than guessing that might give suggestion to the nurse.

The percentage of the performance of baby cry recognition could also be increased by including the sensors. Figure 9 shows the sample of the sensor’s records from the incubator. The humidity was in % as

a relative value compared with saturated vapour density. Meanwhile, the temperature was in Celsius. Finally, MQ-135 and MQ-3 showed the rated flow of air quality and alcohol-ethanol-smoke, respectively. By combining those sensors with the baby cry interpreter, it is possible to know the baby's overall condition. This will improve the current infant incubator [19].

The main idea behind this is that it will help the nurse in monitoring the baby. Yet, a high-cost efficiency using this IoT technology could be achieved by improving the function of actuators inside the incubator. The system will also help the technician in maintaining the incubator. A historical record of the baby cry and the sensors data will give an overall report of the baby condition and its environment.

4. Conclusion

From this system, it is possible to find out the baby voice classification by using machine learning. Our test with the open voice database was possible to get 85% of accuracy. This result was limited with the number of non-hungry group of data which might be the reality of babies at that age range. Meanwhile, the incubator will be used a lot in range ages less than a month. For the sensor, as shown, it showed the possibility of a further process to improve the overall report of baby condition. However, it will need a further clinical trial to find out air condition inside the chamber of the infant incubator.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Erwin Sutanto is the first author and correspondence for this work and article. Fahmi is assisting in the analysis of baby voices, and Wervyan Shalannanda is responsible for the IoT System. Meanwhile, Arga Aridarma is taking care of the infant incubator from concept until the product.

Acknowledgments

The authors would like to thank the Ministry of Research and Technology (Kemenristek/BRIN) Government of Indonesia who supported this work by the Program Penelitian Kolaborasi Indonesia (PPKI) 2020 fund. The scheme was a collaboration between the Universitas Airlangga, Institut Technology of Bandung, and Universitas Sumatera Utara.

Appendix

π	Mathematical Constant of 3.14159
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DBL	Dunstan Baby Language
FN	False Negative
FP	False Positive
IoT	Internet of Things
J	Imaginary unit
N	Total of Negative group
P	Total of Positive group
TP	True Positive
TN	True Negative

References

- [1] N. Sharma, I. Fierens, S. Mohinuddin, N. Ratnavel, S. T. Kempley, and P. Sakhuja, "Servo-controlled thermoregulation in extreme preterm and extremely low birth weight infants during neonatal transport", *Archives of Disease in Childhood-Fetal and Neonatal Edition*, Vol. 105, No. 1, pp. 113–114, 2020.
- [2] A. R. Shabaan, S. M. El-Metwally, M. M. Farghaly, and A. A. Sharawi, "Pid and fuzzy logic optimized control for temperature in infant incubators", in *2013 5th International Conf. on Modelling, Identification and Control (ICMIC)*, pp. 53–59, IEEE, 2013.
- [3] S. K. Singla and V. Singh, "Design of a microcontroller based temperature and humidity controller for infant incubator", *Journal of Medical Imaging and Health Informatics*, Vol. 5, No. 4, pp. 704–708, 2015.
- [4] A. Lahana, S. Delanaud, R. Erbani, A. Glusko-Charlet, E. Durand, E. Haraux, L. Ghyselen, J.-P. Libert, and P. Tourneux, "Warming the premature infant in the delivery room: Quantification of the risk of hyperthermia", *Medical Engineering & Physics*, Vol. 59, pp. 70–74, 2018.
- [5] J. H. Hess, "An electric-heated water-jacketed infant incubator and bed: For use in the care of premature and poorly nourished infants", *Journal of the American Medical Association*, Vol. 64, No. 13, pp. 1068–1069, 1915.
- [6] J. de Araujo, J. de Menezes, A. Moura de Albuquerque, O. da Mota Almeida, and F. Ugulino de Araujo, "Assessment and certification of neonatal incubator sensors through an inferential neural network", *Sensors*, Vol. 13, No. 11, pp. 15613–15632, 2013.
- [7] E. Sutanto, K. AIN, M. AZIZ, and G. ESCRIVA-ESCRIV A, "A study on dc limit parameters in rcd operation using capacitor",

- Journal of Engineering Science and Technology Review*, Vol. 12, No. 4, pp. 7–14, 2019.
- [8] E. Ozdemirci, M. O. Yatak, F. Duran, and M. R. Canal, “Reliability assessments of infant incubator and the analyzer,” *Gazi University Journal of Science*, Vol. 27, No. 4, 2014.
- [9] A. Srinivasan, “IoT cloud based real time automobile monitoring system”, In: *Proc. of 2018 3rd IEEE International Conf. on Intelligent Transportation Engineering (ICITE)*, pp. 231–235, IEEE, 2018.
- [10] S. Angadi and V. S. Reddy, “Hybrid deep network scheme for emotion recognition in speech”, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 3, pp. 59–67, 2019.
- [11] S. Choi, S. Yun, and B. Ahn, “Implementation of automated baby monitoring: Ccbebe”, *Sustainability*, Vol. 12, No. 6, p. 2513, 2020.
- [12] K. Sharma, C. Gupta, and S. Gupta, “Infant weeping calls decoder using statistical feature extraction and gaussian mixture models”, In: *Proc. of 2019 10th International Conf. on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, 2019.
- [13] A. Hamdy, P. K. Vedula, and M. V. J. Konduru, “Audio separation and isolation: A deep neural network approach”.
- [14] G. Veres, “donateacry-corporus: Infant cry audio corpus repository.” <https://github.com/gveres/donateacry-corporus>, 2019.
- [15] D. Anggraeni, W. Sanjaya, M. Solih, and M. Munawwaroh, “The implementation of speech recognition using mel-frequency cepstrum coefficients (mfcc) and support vector machine (svm) method based on python to control robot arm”, In: *Proc. of Annual Applied Science and Engineering Conf.*, Vol. 2, pp. 1–9, 2018.
- [16] J. Lyons, D. Y.-B. Wang, Gianluca, H. Shteingart, E. Mavrincac, Y. Gaurkar, W. Watcharawisetkul, S. Birch, L. Zhihe, J. Hlzl, J. Lesinskis, H. Almr, C. Lord, and A. Stark, “jameslyons/python speech features: _ release v0.6.1”, 2020.
- [17] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems”, 2015. Software available from tensorflow.org.
- [18] M. Ounsted and G. Sleight, “The infant’s self-regulation of food intake and weight gain: difference in metabolic balance after growth constraint or acceleration in utero”, *the Lancet*, Vol. 305, No. 7922, pp. 1393–1397, 1975.
- [19] W. Shalannanda, I. Zakia, E. Sutanto, and F. Fahmi, “Design of hardware module of IoT-based infant incubator monitoring system”, In: *Proc. of 2020 6th International Conf. on Wireless and Telematics (ICWT)*, pp. 1–6, IEEE, 2020.