The Study of Learning System for Infant Cry Classification Using Discrete Wavelet Transform and Extreme Machine Learning

Anyawee Chaiwachiragompol, Nattawoot Suwannata*

Faculty of Engineering, Mahasarakham University, Maha Sarakham 44150, Thailand

Corresponding Author Email: nattawoot.s@msu.ac.th

ABSTRACT

The learning system of infant cry is presented. This system consists of characteristics attraction technique and classification technique. The characteristics attraction of infant cry are based on Discrete Wavelet Transform (DWT) methods. Whilst the sound classification of coefficients characteristics uses Single Layer Neural Feed Forward (SLNF) as an Extreme Learning Machine (ELM). The Dunstan Baby Language (DBL) is the sound database for the proposed system. The sound database was collected from infants between birth and 6 months of age. Where the baby language groups are categorized into 5 types: "Eh", "Eairh", "Neh", "Heh" and "Owh", respectively. The accuracy of sound classification was designated at the number of hidden nodes of 10 – 50 with a training and testing ratio of 70/30. The suitable results are based on the number of epochs, accuracy and performances. The results show that the average accuracy of all discrete wavelet functions on the baby language are over 80%. The average performance of Sym2 is suitable for all baby language groups. Moreover, the average number of epochs of Bior3.1 is suitable for all baby language groups.

1. INTRODUCTION

In 1998, Priscilla Dunstan, in collaboration with research organizations from Australia and the US, established an organization called "Dunstan Baby Language (DBL) [1], which studies babies' behaviors and development. In her studies, in which over 1,000 babies ranging from newborns to 6-month-old babies in 20 countries all over the world were involved, suggest that babies produce 5 universal sounds: Heh, Eh, Owh, Heh, and Eairh.

"Heh" is the sound babies produce most. The sound suggests that the baby is feeling uncomfortable, which may be due to wetness, heat, or sweat. The mother may need to give the child a bath or talcum powder to prevent body moisture and rashes. "Eh" indicates that the baby is experiencing low gas or a stomach upset. The mother may carry the baby facing out on her shoulder and give it gentle pats on its back. Alternatively, the mother may put the baby facing out in a sitting position on her laps, lean the baby slightly forward and give it gentle pats on the back. "Owh" is uttered when the baby is feeling sleepy or in need of a sleep. "Inneh" of "Neh" indicates thirst or hunger. "Eairh" is used to communicate that the baby needs burping due to congested air in the chest or stomach or the show the feelings of discomfort [2]. However, these five sounds can be understood only by baby specialists. For people who are not experienced in classifying babies' universal sounds may take a long time in trials and errors to understand infants' behaviors by listening and properly responding to their cries from the beginning to the end of the cry. Nevertheless, this requires a long range of time to master.

A number of previous studies on the classification of baby cries for caregiver's prompt responses have made a platform for further research on baby behavior development. In previous studies on baby cry classification, cry analysis and processing were performed through voice characterization. The characterized voices were progressed to voice characteristic extraction using artificial neuron network, that is the characteristic extraction process for significant coefficients of baby cry through Mel Frequency Cepstral Coefficients (MFCC). Then the analyzed voices through MFCC were analyzed for accuracy coefficients using Support Vector Machines (SVM) [3]. Later, studies on baby cry classification for detection of parasitic disease, using artificial neuro networks, were carried out using wavelet transform (WT) for voice extraction of babies with a normal physical condition and babies with parasites with an accuracy of 99% [4]. After that, Hariharan M. employed voice frequency analysis and regression analysis to classify of parasite diseases in babies [5]. In 2012, Hariharan compared two techniques: Multilayer Perceptron (MLP) Time-Delay Neural Networks for accuracy in cry-based classification of parasitic diseases in infants [6]. Similarly, in 2015 Alejandro employed fuzzy technique in categorizing baby cry using MFCC characteristic extraction [7]. However, there is no learning systems applied to special sound especially, the baby language groups such as "Eh" "Eairh" "Neh" "Heh" and "Owh".

This study found that sound characterization by MFCC and Discrete Wavelet Transform (DWT) methods is popular because they are an improved technique from cepstrum by adjusting the scale of the spectrum to the appropriate scale. The research has focused on DWT feature extraction in the preprocessing process because the characteristics of the infant's sound are different. Moreover, the DWT also reduces the sample count from the original audio signal. This makes the learning process faster while the learning accuracy is still acceptable.

In this paper, the DWT and Single Layer Neural Feed Forward (SLNF) are applied for learning system of infant cry. Where this system can be a guideline for further research on the determination of optimal baby cry identification with high precision. In addition, it can provide profound insight regarding the characterization of baby behaviors through their cry.

2. RELATED WORKS

2.1 Theory/calculation

2.1.1 Characteristic extraction of wave signals

The infant's cry database of Dunstan baby language is used as the tested data. The DBL collects information from kids between birth and 6-month-old of age in 20 countries, and a total of 30 nationalities. The DWT with Harr, Coif1, Saym2, Db2, and Bior3.1 mother wavelets are used to extract features of the infant cry signal. The data from DWT is classified by Extreme Learning Machine (ELM) at number hidden layers of 10, 20, 30, 40 and 50 nodes. The classification type of infant's sound is "Eairh", "Eh", "Heh", "Neh" and "Owh" [8]. Knowing the meaning, and how to recognize the cry, will respond to the baby's needs correctly. In which Thai babies from birth to 6 months will communicate in the main international languages, which are the sounds that occur on a daily basis, a total of 5 words, namely: [2] "Heh" is the most common sound a child sings to tell a parent or guardian that the baby is feeling unwell. Which is caused by wetness, heat, sticky, stuffy, making the baby's skin wet easily, causing skin irritation. The caregiver or guardian can take care of it by bathing and applying baby powder. To prevent wet and also prevent rashes. An example of a baby's crying signal is shown in Figure 1.
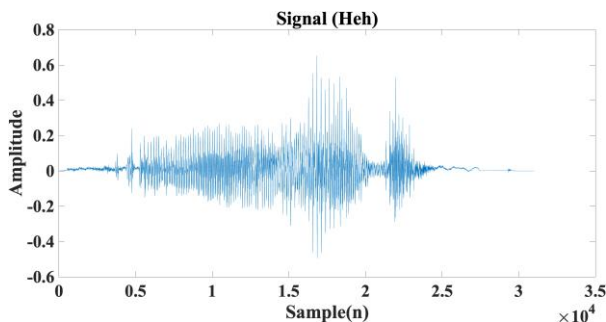

Figure 1. Example of normalized infant crying "Heh" signal

"Eh" tells about the symptoms that the baby has fainting in the stomach. The caregiver or guardian should take care of the child by holding the child up to support the child's head over his shoulder. Face the caregiver and gently pat the baby's back, or perhaps hold the baby up in a sitting position. Lean the baby forward slightly and gently pat the baby's back. An example of a baby's crying signal is shown in Figure 2.

"Owh" describes how sleepy the baby is and wants to rest. Babies usually have the feeling of being hugged by their caregivers or parents. In another way, babies need their caregivers to hug. How to solve the problem when the baby is sleepy, let the caregiver sit or stand holding the baby on the chest and rock it slowly and steadily. An example of a baby's crying signal is shown in Figure 3.
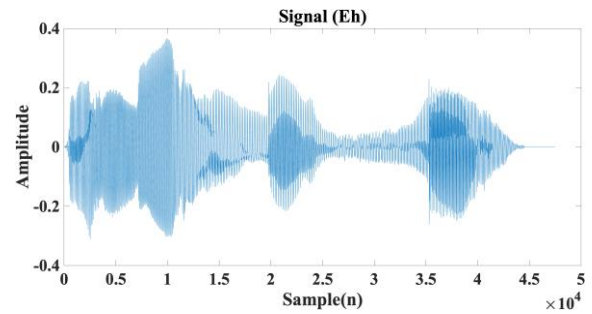

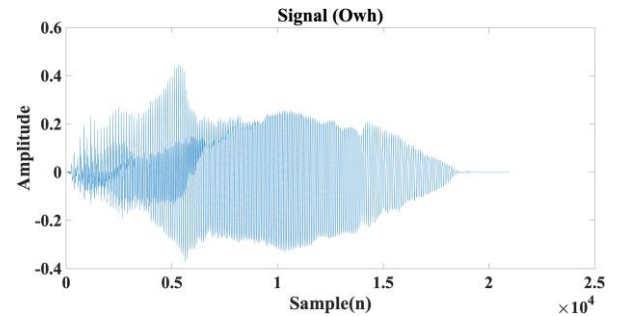**Figure 2.** Example of normalized infant crying "Eh" signal

Figure 3. Example of normalized infant crying "Owh" signal

"Neh" describes the symptoms that the baby is thirsty or thirsty. I want mothers to feed milk and water to feel full. By the nature of the sound, it is like a nosebleed, the caregiver sits and hugs the baby while feeding the baby, using a pillow to support the baby to keep the baby warm. An example of a baby's crying signal is shown in Figure 4.
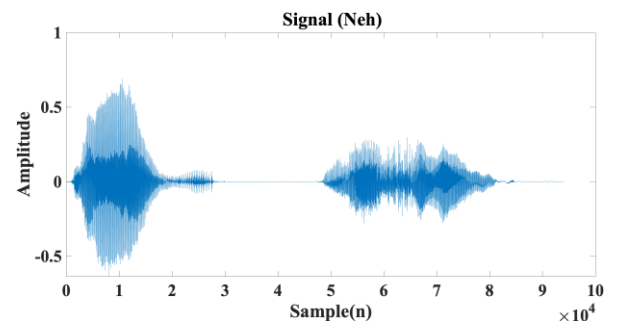

Figure 4. Example of normalized infant crying "Neh" signal

"Eairh" describes a baby with bloating or abdominal pain that resembles a fuss. from the initial symptoms of burping with intestinal disorders, the baby will feel uncomfortable. The emission characteristic is to open the mouth wide. The baby will emit a sound from the baby's belly. The babysitter helps the baby by holding the baby over his shoulder, gently stroking or patting the baby's back. An example of a baby's crying signal is shown in Figure 5.

The baby cry signal is taken from a dataset through an audio signal normalization process to convert the audio signal into a digital signal, extracting the features of cried. The sampling rate at 44 hertz was used, each sampling had a resolution of 16 bits, a sample size of the sound used to describe the amount of data stored in a stereo. In the discrete wavelet transform (DWT) conversion process, the signal is passed through a low-pass filter and a high-pass filter to separate the signal into low-pass segments where the low-pass filter is the scaling function, and the high-pass filter is the wavelet function. The result will be

half the frequency of the original signal. After that, 25 infant cry signals were extracted from the DWT method.

Baby cries are proceeded for characteristic extraction for significant wave signals using discrete wavelet transform (DWT) developed from Theory of Wavelet Transforms, which will be described below.
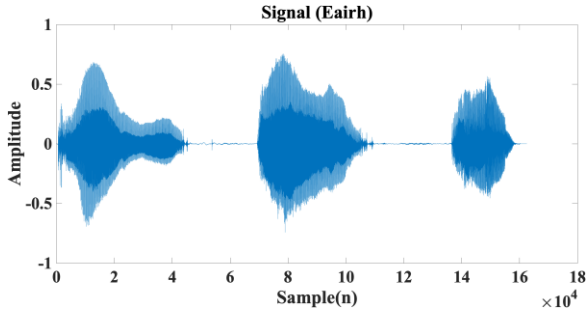

Figure 5. Example of normalized infant crying "Eairh" signal

2.1.2 Wavelet transforms

Wavelet transform is a method used to analyze discrete signals and present them in the time domain, while the Fourier transform presents the signal in the frequency domain and only tells what frequencies are available in the signals. To overcome the shortcomings, Short-Time Fourier Transform (STFT) was developed. The STFT can know what frequency occurs at any given time by using the wavelet function for comparison with the signal. However, STFT is limited to a trade-off between time and frequency resolution in STFT, so wavelet transform was developed. Wavelet transformations are scaling analysis and study of composition and resolution a signal for frequencies by decomposing and shifting factors of a single wavelet function. Because Wavelet Theory is the mathematical that explains a signal machine or a physical process, which assumes that any given signals or machines are composed of sets of small specific signals called "wavelets" as shown in Figure 6 [9].

Figure 6. Wavelet characteristics

To explain a signal or system, a string of wavelets from same function are composed to form a mother wavelet. Individual wavelets are later scaled and translated, and always normalized with $1/\sqrt{a}$. Therefore, given $g(t)$ the mother wavelet, a general wavelet equation at any (a, b) related to the mother wavelet is defined as:

$$g_{a,b}(t) = \frac{1}{\sqrt{a}} g(\frac{t-b}{a})$$ (1)

A vector space is a signal space formed by the aggregation of sub-signals known as the Basis Function, given v being a vector space and j being the degree of resolution. The number of basis functions compiling a function is called scaling function, it can be called the scaling function [$\Phi(t)$] which refines resolution level with frequency bands. In other words,

for low frequency range, the resolution is low. Likewise, for high frequency range, the resolution is high. the value of the frequency, if the frequency range is low, the resolution level is low. Reduction in resolution degree by one level results in twice lower level of basis functions. Using multi-resolution signal analysis, $f(t)$ at any j vector spaces can be defined as in Eq. (2) [10]:

$$f_j(t) = \sum_k C_k^j \, \Phi_k^j(t)$$ (2)

where, C_k^j is scaling coefficient obtained from resolution refinement j at any k positions. j and k are resolution level and location for analysis of the signal, respectively. $\Phi_k^j(t)$ is the relationship of scaling functions in any vector space and is expressed in Eq. (3):

$$\Phi_k^j(t) = 2^{\frac{-j}{2}} \Phi(2^j t - k); \; j, k \in z$$ (3)

where, z is set of integers. According to Eq. (2), the source signal $f(t)$ can be determined by multiplying the scaling coefficient C_k^j with k and j. In an analysis of lower resolution, using such equation may result in signal loss in another vector space called wavelet vector space W^j composed of a basis function called wavelet function $\varphi(t)$ as suggested in the following Eq. (4):

$$\Psi_k^j(t) = \frac{1}{2^{j/2}} \Psi(2^t t - k)$$ (4)

Given $g_j(t)$ is a signal generated from $\Psi_k^j(t)$ under the same vector space, when combined into any signals, the signal generated is following:

$$g_j(t) = \sum_k d_k^j \, \Psi_{kj}(t)$$ (5)

Assuming that d_k^j represents scaling coefficient obtained from resolution refinement j at any k positions to generate $g(t)$, therefore, considering the relationship $V^{j-1} \oplus W^{j-1} = V^j$. Therefore, the multi-resolution signal analysis, $f(t)$ at any j vector spaces can be defined as:

$$f_j(t) = f_{j-1}(t) + g_{j-1}(t)$$ (6)

If $f(t) \in V^j$ is decomposed so that $f(t)$ has lower resolution as a result of $V^{j-1} \oplus W^{j-1} = V^j$, while V^j can be broken down until $j=0$, which yields Eq. (7):

$$V^j = V^0 \oplus W^0 \oplus W^1 \oplus ..... \oplus W^{j-1}$$ (7)

Similarly, $f_j = f_{j-1} + g_{j-1}$ can also be decomposed to f_{j-1} and g_{j-1}. When plotted on a scatter diagram of different resolution sequences, $f(t)$ can thus be expressed as a resolution adjustment function and a wavelet function as in Eqns. (8) and (9):

$$f(t) = f_0 + g_0 + g_1 + g_2 + ... + g_{j-1}$$ (8)

$$f(t) = \sum_k C_k^o \, \Phi_k^o(t) + \sum_{j=0}^{j-1} \sum_k d_k^j \, \Psi_k^j(t)$$ (9)

Therefore, wavelet coefficient can be examined from Eqns. (10) and (11):

$$C_k^j = \langle f(t), \Phi_k^j(t) \rangle \qquad (10)$$

$$d_k^j = \langle f(t), \Psi_k^j(t) \rangle \qquad (11)$$

where, $\langle , \rangle$ denotes dot product. The $f(t)$ is decomposed to have resolution j. The coefficient C_k^j is then broken down to C_k^0 in the sets of d_k^{j-1}, ..., d_k^1, d_k^0 at different resolution levels. The $C_j(m)$ can be determined through circular convolution between mother wavelet and $\phi_{j,m}(t)$. Consequently, the $C_j(m)$ can be determined through circular convolution between mother wavelet and mother wavelet and $\Psi_{j,m}(t)$.

The input signals are classified into 2 levels of coefficients: approximation coefficient and detail coefficient obtained via circular convolution between source the signal element with low frequency and high frequency filter, with the sampling rates reduces by two times, where n represents level of resolution in integer, starting from 1, 2, …, ∞. high-pass and low-pass filters in Discrete wavelet transform can be operated using the following equations:

$$Y_{high}(k) = \sum_n x(n)h(2k - n) \qquad (12)$$

$$Y_{low}(k) = \sum_n x(n)g(2k - n) \qquad (13)$$

The $Y_{high}(k)$ and $Y_{low}(k)$ are the results from high frequency filter and low frequency filter activated by h and g, respectively [11]. All signals are then extracted for cry classification using 5 scaling functions of wavelet transform, namely, Haar, Db2, Coif1, Sym2, and Bior3.1.

2.2 Wavelet function

Typically, the wavelet function consists of Harr wavelet, Daubechies wavelet, Symlets wavelet, Coiflets wavelet and Biorthogonal wavelet, respectively. Firstly, Haar wavelet is discontinuous and resembles a step function. It represents the same wavelet as Daubechies db1. Figure 7 shows Haar wavelet [12].
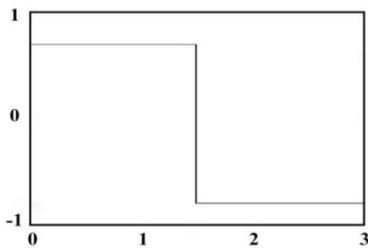

Figure 7. Haar wavelet

Secondly, this family was introduced by Ingrid Daubechies, who has pioneered much research in the wavelet transform domain. The surname of this family is 'db' and there are nine members belonging to this family, from 'db2' to 'db10'. The number next to the surname represents the order N, so the general form for this family is 'dbN'. The general characteristics of this family are orthogonal, compactly

supported, with highest number of vanishing moments N and almost all members are unsymmetrical. In case of low-level distortion this family gives better accuracy when using mother wavelet with high order N. Figure 8 shows db2 wavelet [12].
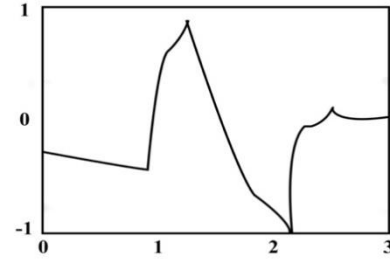

Figure 8. Db2 wavelet [13]

Thirdly, in order to modify the symmetry property of the 'db' family, Ingrid Daubechies invented this family. This family is almost symmetrical, its surname is 'sym' and there are seven members belonging to this family, from 'sym2' to 'sym8'. This family is orthogonal, compactly supported in time and has N vanishing moments. Figure 9 shows sym2 wavelet [12].

Fourthly, at the request of R. Coifman, Ingrid Daubechies built this family. The surname of this family is 'coif' and there are five members belonging to this family, from 'coif1' to 'coif5'. This family is characterized by its highest number of vanishing moments, which is 2N. It is compactly supported and orthogonal but is near to symmetry. In case of high-level distortion this family gives better accuracy when using mother wavelet with low order N. Figure 10 shows coif1 wavelet [12].

Finally, biorthogonal wavelet is the family of wavelets exhibits the property of linear phase, which is needed for signal and image reconstruction. By using two wavelets, one for decomposition (on the left side) and the other for reconstruction (on the right side) instead of the same single one, interesting properties are derived. Figure 11 shows bior3.1 wavelet [12].
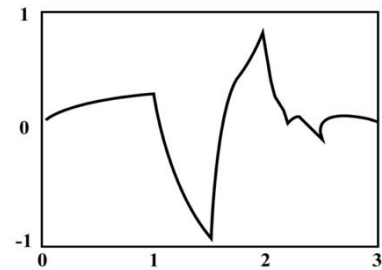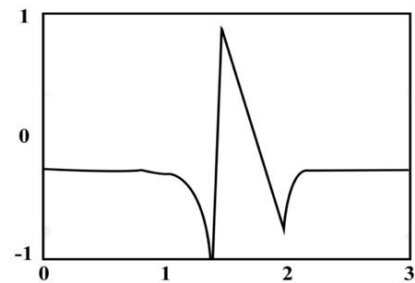

Figure 9. Sym2 wavelet [13]
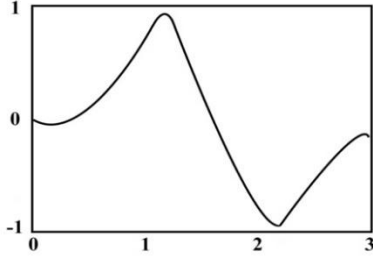

Figure 10. Coif1 wavelet [13]

Figure 11. Bior3.1 wavelet [13]

2.3 Data preparation

The data obtained from feature extraction in 2.3 yield a matrix form of coefficients before being classified, for accuracy of analysis, all sets of data will be cross-checked and arranged in uniform data set [3] as expressed in Eq. (14):

$$data' = \frac{data - Min}{Max - Min} \qquad (14)$$

The data include the characteristic information of each cell, and max and min value represents the maximum and minimum values of each characteristic of each data set. The experiment has a proportion of 70% training and 30% testing, of which the data in each class are equally divided. The data to be operated need to be classified in terms of normal cell and abnormal cell.

2.4 Artificial neural network machine

Extreme Learning Machine (ELM) was first introduced by Huang et al. [5] which features a technique called Single Layer Feed Forward (SLFN) comprising hidden layer and output layer as in Example [2]. Figure 12 Show the structure of SLFN.
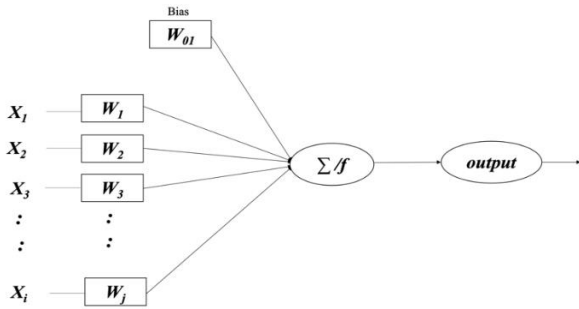

Figure 12. The SLFN structure [2]

The data sets for SLFN machine are comprised of N data, and each pair or reports is represented by (x_i, t_i). As stated by [2], SLFN consist of $\tilde{N}$ hidden nodes.

SLFN function in extreme process is started by sampling weight and deviation of node j calculated by $f(W_j X_i + b_i)$, where f represents nonlinear activation function on the output. The normal equation derived from 15 can be expressed as follows [5]:

$$H\beta = T \qquad (15)$$

where, $H = \{h_{ij}\}$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, \tilde{n}$ are result matrices from the hidden layers, and T represents target results. Solution for weighted equation of β layer can be obtained from Eq. (16):

$$\overset{\wedge}{\beta} = H^+ T \qquad (16)$$

where, H^+ represents the Moore-Penrose pseudoinverse of matrix, in which $\overset{\wedge}{\beta}$ denotes the minimum norm of Eq. (17):

$$\left\| H\overset{\wedge}{\beta} - T \right\| = \left\| HH^+ T - T \right\| = \min_{\beta} \left\| H\beta - T \right\| \qquad (17)$$

The Moore-Penrose pseudoinverse of matrix [4] is calculated using SVD technique [7].

2.5 Performance verification

When the model for machine learning is obtained, model testing to compare model performance is determined by the validity of the training and testing dataset. Exhaustive cross-validation methods are cross-validation methods which learn and test on all possible ways to divide the original sample into a training and a validation set. The Cross Validation simulation test has the following workflow.
 - Divide the learning data into k equal sets.
 - Use the rest of the data (k- sets) to create the model.
 - Collect 1set of divided data to Evaluate to find the error from model for machine learning by the Mean Squared Error (MSE) as the Eq. (18):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \qquad (18)$$

where, n is the number of samples of the data set, y_i is the actual result and $\tilde{y}_i$ is the result of the presision in the model.

Iterate until all parts of the data are tested. To measure the accuracy of the present study, target information and ELM data are calculated for accuracy percentage as shown in Eq. (19):

$$accuracy(Out) = \sum_{i=1}^{Out} \frac{Out_i}{T}, Out_i \in T \qquad (19)$$

where, *accuracy (Out)* represents accuracy percentage, Out_i is number of correctly classified data, and T is the total number of data. The accuracy value of the whole data set can be obtained as the Eq. (20):

$$overall\ accuracy = Average(\sum_{i=1}^{k} Accuracy_i) \qquad (20)$$

3. THE PROPOSED METHOD

Process of voice characteristic extraction for neuro network classification proposed in the present study is shown in Figure 13.

The input signals (infants cries) are processed through frequency transformation, and analyzed for characteristic extraction using discrete wavelet transform function to obtain characteristic coefficients. The data are then compared and arranged in data sets for evaluation of classification accuracy.

The details of process sequence are shown as follows:
1) In the signal normalization process, the amplitude of the

sound is normalized.

2) In the preprocessing process, the DWT separates the normalized signal into an approximation coefficient (or low frequency component, $g[n]$) and a detail coefficient (or high frequency component, $h[n]$). The number of samples of low frequency components and high frequency components are reduced by two times, which is a signal compression process. In this research, the signal $g[n]$ is cascaded by the next 4 levels of DWT. At the 5th DWT level, the number of samples of the $g[n]$ signal is reduced by 32 times as shown in Figure 14. This compression process improves the learning time of the neural network (NN).

3) In the NN classification process, the compressed signal $g[n]$ is classified by ELM model into five sounds: Heh, Eh, Owh, Heh, and Eairh.
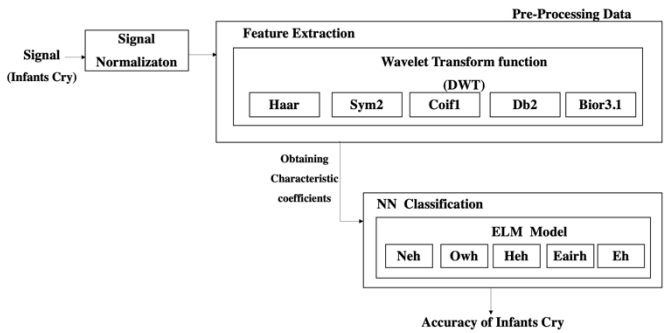

Figure 13. Process of character selection for voice classification using neutron network
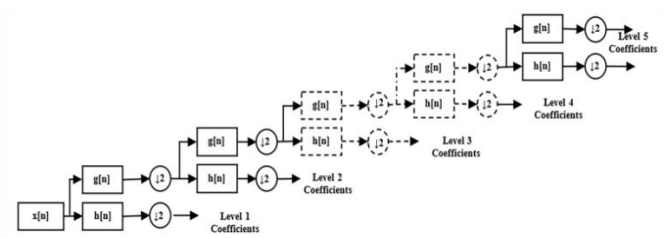

Figure 14. Frequency filter level design structure

The overall flow of the wavelet decomposition process is conducted in many steps as shown in Figure 14. Firstly, the raw infant crying signal $x[n]$ is decomposed into several levels of frequency bands, then the approximate and detail coefficients in each level with standard deviation values for approximate and detail coefficients are found, finally [14].

4. THE RESULTS AND DISCUSSION

4.1 Results

Baby cries are proceeded for characteristic extraction for significant wave signals using DWT form Harr, Coif1, Saym2, Db2, and Bior3.1 wavelet function. It has been found that the scaling function is suitable for compressing sample signals. In the 5-level pass filter is shown in Figure 14 Frequency filter level design structure. An example of a baby crying signal when separating frequency between low-pass filters and high-pass filters is shown in Figure 15.

The coefficient of the extracted high-pass and low-pass filters is used to extract the significant coefficients of that signal from the convolution between the master signal and the

high-pass filter, respectively. When the coefficients for extracting the characteristics of the signal were obtained, the cries coefficient was then subjected to the data preparation process before being introduced into the ELM neural network model (Figure 16). The preparation of the coefficients obtained from the extraction of the cry adjust the values to range -1 to 1.
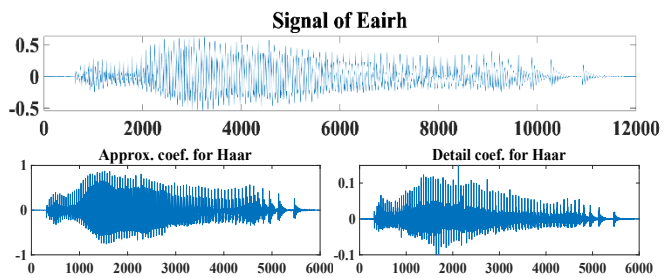

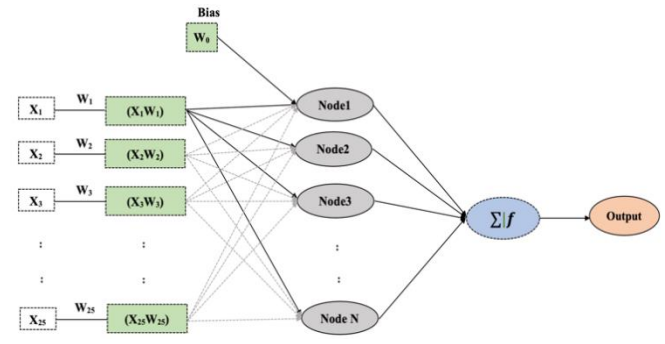**Figure 15.** Examples of signals in the low-high frequency segments of the Haar function in level 2

Figure 16. The ELM neural network model

4.2 Discussion

The ELM was trained using 60 input audio files extracted from official Dunstan 207 labeled videos recorded in Australia: 15 recordings for "Neh", 78 recordings for "Eh", 16 recordings for "Owh", 28 recordings for "Eairh" and 70 recordings for "Heh".

Feature extractions of coefficients from baby cries were performed using discrete wavelet transform. From 25 sounds tested, a total of 5 wavelet functions: Harr, Coif1, Saym2, Db2, and Bior3.1, were involved in the character extractions. Each of the Haar wavelet functions produced 5,887 samples, coif1 produced 5,896 samples, sym2, db2 and bior3.1 wavelet functions produced 5,885 samples each.

Table 1. Summary of average accuracy rate of cry classification at 10–50 hidden nodes

Cry classification	Nodes of Average Accuracy %				
	10	20	30	40	50
Eh	89.47	94.74	84.21	94.74	94.74
Eairh	78.95	73.69	94.74	94.74	94.74
Neh	78.95	89.47	84.21	84.21	84.21
Heh	94.74	94.74	84.21	84.21	89.47
Owh	89.47	94.74	94.74	84.21	94.74
Average	86.32	89.48	88.42	88.42	91.58

Data obtained from frequency filters in Level 5 were imported to extreme learning machine for baby cry classification, divided into 5 audio groups: Neh, Owh, Heh,

Eairh, and Eh. The characterized sounds were compared for the computation for classification accuracy designated in 10, 20, 30, 40, and 50 hidden nodes, respectively with the training: testing rate of 70:30 [15]. The results are as shown in Table 1.

Table 1 summarizes average accuracy rates of classification at 10 – 50 nodes of 5 classes of infant cry from 5 wavelet functions. "Eh" achieves the highest average accuracy rate of 94.74% at 20, 40, and 50 hidden nodes. "Eairh" receives an average accuracy rate of 94.74% at 30, 40, and 50 hidden

nodes. "Neh" gets an average accuracy rate of 89.47% at 20 hidden nodes, while "Heh" obtains the best average accuracy rate of 94.74% at 10 and 20 hidden nodes. Finally, "Owh" has the highest average accuracy rate of 94.74% at 20, 30 and 50 nodes. It is remarkable that neuron network extreme learning machine at 50 hidden nodes yield the highest overall average accuracy rate of cry classification. The results suggest that the average accuracy of all 5 cry classifications from 5 wavelet functions are over 80% at 30, 40, and 50 hidden nodes.

Table 2. Summary of performance and epoch numbers of the learning machine at 10 – 50 hidden nodes

Node-Model	Infant Cry of classification model									
	Eh		Eairh		Neh		Heh		Owh	
	Performance	Epoch	Performance	Epoch	Performance	Epoch	Performance	Epoch	Performance	Epoch
10	1.077	4	0.605	6	0.074	2	0.517	5	0.637	7
20	0.689	9	0.615	2	0.082	7	0.601	6	0.377	12
30	0.465	5	0.650	5	0.054	4	0.211	10	0.242	13
40	0.212	10	0.174	8	0.127	4	0.138	2	**0.851**	**3**
50	0.055	13	0.212	10	0.044	5	0.173	1	0.574	9

Table 3. Summary of baby cry classification of extreme learning machine of Haar wavelet

Cry classification	Accuracy	Performance	Epoch
Eh	**91.58**	**0.50**	**8**
Eairh	87.37	0.45	6
Neh	84.21	0.08	4
Heh	89.47	0.33	5
Owh	91.58	0.54	9
Average	88.84	0.38	6

As shown in Table 3, the average accuracy in cry classification of extreme learning machine of Harr function on "Eh" produces the highest accuracy of 91.58% in 8 epochs, which has a lower epoch number comparing to "Owh" sound.

Table 2 shows a summary of learning machine performance and epoch number at 10 to 50 hidden nodes with a classified of 5 different types of infant cry. "Eh" yields a performance of 1.077 at 10 hidden nodes, while "Eairh" gives a performance of 0.650 at 30 hidden nodes, and "Heh" has a performance of 0.851 at 20 hidden nodes. Furthermore, "Neh" has a performance of 0.127 at 40 hidden nodes, and "Owh" obtains a performance of 0.851 at 40 hidden nodes. The results also suggest that extreme learning machine at hidden layer 40 gives the highest performance of learning machine of infant cry classification.

Table 4. Summary of baby cry classification of extreme learning machine of Db2 wavelet

Cry classification	Accuracy	Performance	Epoch
Eh	84.21	0.26	9
Eairh	84.21	0.49	14
Neh	88.42	0.23	7
Heh	89.47	0.39	7
Owh	91.58	0.37	9
Average	87.57	0.35	9

The results from Table 1 show that all the 5 wavelet functions produce significant abilities in extracting features of infant cry using DWT to determine significant coefficients of the analyzed signals. Their classification accuracy levels by types of wavelets are discussed below.

Table 4 presents average accuracy values in cry classification of extreme learning machine using Db2 wavelet.

As we can see, "Owh" yields the highest accuracy 91.58% in 9 epochs.

As shown in Table 5, the highest average accuracy value in cry classification of extreme learning machine using Sym2 wavelet is 87.37% in 6 epochs.

As presented in Table 6, the highest average accuracy value in cry classification of "Heh" sound of extreme learning machine using Coif1 wavelet is 90.53% in 7 epochs.

The results in Table 7 suggest that the average accuracy value in cry classification of extreme learning machine of Bior3.1 wavelet on "Neh" sound produces the highest accuracy of 87.06% in 3 epochs.

Table 5. Summary of baby cry classification of extreme learning machine of Sym2 wavelet

Cry classification	Accuracy	Performance	Epoch
Eh	82.11	0.50	6
Eairh	84.21	0.40	9
Neh	84.21	0.32	8
Heh	83.16	0.37	5
Owh	87.37	0.57	6
Average	84.21	0.43	7

Table 6. Summary of baby cry classification of extreme learning machine of Coif1 wavelet

Cry classification	Accuracy	Performance	Epoch
Eh	84.71	0.60	5
Eairh	80.00	0.46	8
Neh	87.80	0.29	4
Heh	90.53	0.11	7
Owh	78.60	0.55	5
Average	84.32	0.40	6

Table 7. Summary of baby cry classification of extreme learning machine of Bior3.1 wavelet

Cry classification	Accuracy	Performance	Epoch
Eh	83.18	0.41	6
Eairh	84.71	0.63	5
Neh	87.06	0.15	3
Heh	84.71	0.19	3
Owh	81.18	0.64	3
Average	84.17	0.40	4

From the results in Table 3 to Table 7, we found that the average learning accuracy by ELM neural network of all discrete wavelet functions on the baby language are over 80%.

5. CONCLUSIONS

The learning system of infant cry is presented. This system consists of Discrete Wavelet Transform and Single Layer Neural Feed Forward as an Extreme Learning Machine. This learning system is applied to "Dunstan Baby Language (DBL)" which is the special sound database. The average accuracy of all discrete wavelet functions on baby language are over 80%. The average performance of Sym2 is suitable for all baby language groups. Moreover average number of epochs of Bior3.1 is suitable for all baby language groups. The learning system of infant cry using learning machine can be a guideline for further research on determination of optimal baby cry identification with high precision. Also, the results can provide profound insight regarding characterization of baby behaviors through their cry, which will be beneficial to hospitals and nurseries.

ACKNOWLEDGMENT

REFERENCES

[1] Available from: video: The Dunstan Baby Words phonetic descriptors are copyrighted © 2006 Dunstan Baby. http://www.dunstanbaby.com/our-research/.

[2] Renanti, M.D., Buono, A., Kusuma, W.A. (2013). Infant cries identification by using codebook as feature matching, and MFCC as feature extraction. Journal of Theoretical and Applied Information Technology, 56(2): 437-442. http://repository.ipb.ac.id/handle/123456789/76356

[3] Amaro-Camargo, E., Reyes-García, C.A., Arch-Tirado, E., Mandujano-Valdés, M. (2007). Statistical vectors of acoustic features for the automatic classification of infant cry. International Journal of Information Acquisition, 4(4): 347-355. https://doi.org/10.1142/S0219878907001423

[4] Hariharan, M., Yaacob, S., Awang, S.A. (2011). Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. Expert Systems with Applications, 38(12): 15377-15382. https://doi.org/10.1016/j.eswa.2011.06.025

[5] Hariharan, M., Sindhu, R., Yaacob, S. (2012). Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network. Computer Methods and Programs in Biomedicine, 108(2): 559-569. https://doi.org/10.1016/j.cmpb.2011.07.010

[6] Hariharan, M., Saraswathy, J., Sindhu, R., Khairunizam, W., Yaacob, S. (2012). Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks. Expert Systems with Applications, 39(10): 9515-9523. https://doi.org/10.1016/j.eswa.2012.02.102

[7] Rosales-Pérez, A., Reyes-García, C.A., Gonzalez, J.A., Reyes-Galaviz, O.F., Escalante, H.J., Orlandi, S. (2015). Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model. Biomedical Signal Processing and Control, 17: 38-46. https://doi.org/10.1016/j.bspc.2014.10.002

[8] Chaiwachiragompol, A., Suwannata, N. (2016). The features extraction of infants cries by using discrete wavelet transform techniques. Procedia Computer Science, 86: 285-288. https://doi.org/10.1016/j.procs.2016.05.073

[9] Young, R.K. (1993). Wavelet theory and its applications. Boston: Kluwer Academic Publishers, 1-4.

[10] Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine intelligence, 11(7): 674-693. https://doi.org/10.1109/34.192463

[11] Raghuveer, M.R., Bopardikar, A.S. (2002). Wavelet Transforms: Introduction to Theory and Applications, 3rd edn (Singapore: Pearson Education (Singapore)), pp. 25-50.

[12] Chourasia, V.S., Mittra, A.K. (2009). Selection of mother wavelet and denoising algorithm for analysis of foetal phonocardiographic signals. Journal of Medical Engineering & Technology, 33(6): 442-448. https://doi.org/10.1080/03091900902952618

[13] Matlab Wavelet Toolbox User's Guide, Mathwork Incorporation. (2009). http://www.mathworks.com.

[14] Belmahdi, R., Mechta, D., Harous, S. (2021). A survey on various methods and algorithms of scheduling in Fog Computing. Ingénierie des Systèmes d'Information, 26(2): 211-224. https://doi.org/10.18280/isi.260208

[15] Franti, E., Ispas, I., Dascalu, M. (2018). Testing the universal baby language hypothesis - automatic infant speech recognition with CNNs. In 41st International Conference on Telecommunications and Signal Processing TSP 2018, pp. 1-4. https://doi.org/10.1109/TSP.2018.8441412

NOMENCLATURE

Greek symbols

β	Weight value
$\hat{\beta}$	Denotes the minimum norm
Φ	The relationship of scaling functions in any Vector space
ϕ	Convolution between mother wavelet
$\oplus$	Considering the relationship
Ψ	Vector space