International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:7, 2012

A Study on the Differential Diagnostic Model for Newborn Hearing Loss Screening

Chun-Lang, Chang

Abstract**—**According to the statistics, the prevalence of congenital hearing loss in Taiwan is approximately six thousandths; furthermore, one thousandths of infants have severe hearing impairment. Hearing ability during infancy has significant impact in the development of children's oral expressions, language maturity, cognitive performance, education ability and social behaviors in the future. Although most children born with hearing impairment have sensorineural hearing loss, almost every child more or less still retains some residual hearing. If provided with a hearing aid or cochlear implant (a bionic ear) timely in addition to hearing speech training, even severely hearing-impaired children can still learn to talk. On the other hand, those who failed to be diagnosed and thus unable to begin hearing and speech rehabilitations on a timely manner might lose an important opportunity to live a complete and healthy life. Eventually, the lack of hearing and speaking ability will affect the development of both mental and physical functions, intelligence, and social adaptability. Not only will this problem result in an irreparable regret to the hearing-impaired child for the life time, but also create a heavy burden for the family and society. Therefore, it is necessary to establish a set of computer-assisted predictive model that can accurately detect and help diagnose newborn hearing loss so that early interventions can be provided timely to eliminate waste of medical resources. This study uses information from the neonatal database of the case hospital as the subjects, adopting two different analysis methods of using support vector machine (SVM) for model predictions and using logistic regression to conduct factor screening prior to model predictions in SVM to examine the results. The results indicate that prediction accuracy is as high as 96.43% when the factors are screened and selected through logistic regression. Hence, the model constructed in this study will have real help in clinical diagnosis for the physicians and actually beneficial to the early interventions of newborn hearing impairment.

Keywords**—**Data mining, Hearing impairment, Logistic regression analysis, Support vector machines

I. Introduction

HEARING loss can cause insufficiency of abstract thinking, thereby affecting learning ability and achievement; as a result, hearing-impaired children are often misunderstood as a general lack of intelligence. However, the learning outcomes of these children vary depending on the degree of hearing loss, learning motivations, learning environments, the teaching strategy, and the level of intelligence[1]. According to statistics in the literatures, on average five out of a thousand babies have hearing impairment; moreover, at-higher-risk infants and young children have as high as 5-20% rate of hearing loss. Among all children with hearing loss, approximately 50% of them are born with it whereas the other half gradually loses their hearing after birth[2].

About 50% of hearing loss is genetic-related. A hearing impairment can exist in only one ear or in both ears; it can be mild or severe; it can be congenital, late onset, or progressive; it can very likely combine abnormality of other sections to develop into a certain syndrome. Unfortunately, no effective methods in clinical settings can treat these different types of hearing loss; it cannot even be controlled or stopped from getting worse. The only doable way to prevent hearing impairment is to hopefully persuade people of high-risk groups not to have babies through genetic counseling[3].

II. Literature Review

A. Hearing Impairment

Mild hearing loss between 26 and 40 decibels of hearing loss (dB HL): One will be less able to understand speech or become completely nonresponsive to the sound when it is less audible or in a noisy environment; however, when approaching closer to the source of sound or making it louder or more audible, one can clearly hear the speech. Mild hearing loss is difficult to identify but patients can get the most benefit when wearing hearing aids[4],[5].

Hearing loss between 41 and 61 dB HL: One has really slow responses to the sounds in his/her daily life unless the sound is really loud. For people who are deaf at birth, they might not be able to speak or only can say one or two words at the age when they are supposed to speak. Wearing hearing aids might help remedy the situations; however, their hearing and language abilities will only improve when proper hearing and speaking trainings are received[6], [7],[8].

Hearing loss above 91 dB HL: One has almost no responses to any sound but can sense the vibrations caused by sound. Therefore, the deaf could feel when the drums are pounding, and sometimes when an aircraft flew at low altitude, they also responded. Generally speaking the deaf depends on their vision to adapt to the society. Unless going though special trainings in hearing and speaking, they will not be able to learn to speak[9]. Hearing aids have less significant help to the deaf than to patients with mild, moderate, and severe hearing loss, but they can help with the awareness of sound as well as being beneficial in learning to speak and adjusting to the environment[10],[11].

B. Data Mining

Data Mining, in general, can be interpreted as knowledge discovery in the database, KDD. It is to discover certain knowledge from the database and to extract and find the potential useful information hidden in the database to provide references for decision-makers. Its cycling procedures are as listed below:

1. Data Selection: Focusing on the objectives of the study, find the relevant data from the database.
2. Data Integration: Organize and integrate all the data

Chun-Lang Chang is with the Department of Industrial Management at National Formosa University, Taiwan. (e-mail: jcchang@nfu.edu.tw)

collected from the different sources.

3. Data Cleansing: Screen the data to remove the anomalies of a target data set and filter data that is inappropriate or that does not match the requirement.
4. Data Transformation: Reorganize data based on the required method, and then transform data into forms appropriate for mining.
5. Data Mining: Apply AI research techniques on the transformed data to construct a model pattern.
6. Model Evaluation and Interpretation: Use tools to measure and evaluate the accuracy of the model and interpret the results.
7. Knowledge Discovery: Use other related techniques to explain the knowledge acquired through mining and help users understand.

C. Logistic Regression Analysis

With the objective to find the relationship between the response valuables of the class type and a series of explanatory valuables, the biggest difference of logistic regression from general regression analysis lies in the different forms of response variables. Therefore, in terms of its applications, logistic regression also meets the general underlying assumptions in traditional regression analysis, i.e. avoiding the problem of collinearity between explanatory variables, meeting the normal distribution, and avoiding the existence of self-related residuals, etc. The response variables in logistic regression are discretely distributed. When only two or very few classes are involved, logistic regression is the most standard method of analysis. One characteristic is that the dependent variables of logistic regression are binary, such as “disappeared” or “retained”, and there is no assumptions regarding distribution of its independent variables so they may be continuous, discrete, or dummy variables.

D. Support Vector Machine (SVM)

In the early 1990s, Vanik developed support vector machine (SVM) to solve classification problems. Then in 1995, he applied SVM to regression problems. SVM and back propagation network (BPN) has two major differences. First, SVM uses structural risk minimization (SRM) to replace empirical risk minimization (ERM). When training BPN, the only objective is to reduce errors to the lowest; in other words, it only considers ERM. However, SVM, which bases the principle of SRM, not only takes into account of ERM, but also considers model complexity, thereby addressing and avoiding the problem of over fitting. Also, the use of SRM in SVM brings better generalization ability. Second, SVM and BPN have different methods in deciding model structures and weights. For BPN, the weights and model structures are decided and produced using the trial and error method and iterative process, but this process is very time-consuming. To save time wasted during the process of training, Vapnik proposed to change the decision-making process of SVM on parameters and structures into a quadratic programming problem and to be able to use standard algorithms for quick solutions in the meantime.

III. Methods

A. Research Structure

The model construction analysis in this study can be divided into three main stages. The first stage is data collection and pre-processing, conducting normalization and coding on the data in the database. The second stage involves SVM computing analysis, logistic regression combined with SVM, and then compares which analysis method has results with higher accuracy. The third stage focuses on the comparisons and explanations of the results from each model.

1. Collecting medical history data Physiological test data of the newborns during the pregnancy, at birth, and after birth will be obtained by the case hospital.
2. Screening and organizing variables Physiology-related risk factors of newborns diagnosed with hearing loss based on professional physicians’ recommendations and relevant literatures.
3. Processing dataTransform data into executable format based on previous literatures and doctors’ suggestions and conduct normalization to reduce impurities and incomplete information that might have affected accuracy of the predictive model in this study.
4. Training samples and testing numbers of sample clusters Utilize K-fold Cross-Validation to divide samples into K sub-set clusters. Conduct normalizations and turn data into the compatible format for the decision tree algorithm to calculate and analyze accuracy of the models.
5. Constructing models Build models using the impact factors and the analysis goals. Model I: Input variables that impact analysis of causes to SVM to operate and analyze the results. Model II: Input variable that impact analysis of causes to logistic regression to filter the significant variables. Then input the filtered variables into SVM to operate and analyze the results.
6. Analyzing data Use the construction rules on the test data to classify information.
7. Comparing and discussing results Analyze, compare and explain the accuracy of each experimental model.

B. Data Collecting

Subjects in this study are data from the neonatal database of a certain case hospital, taking the neonatal data of babies who were born in the last five years in that hospital. Using these neonatal data with normal hearing and that with confirmed hearing loss, a total of 600 pieces; after deleting incomplete and incorrect data, 563 pieces retained for use in the research.

C. Data Processing

Factors influencing newborn hearing loss are organized as illustrated in Table I. In all, 17 variables, length of stay in the intensive care unit, respiratory distress syndrome, retrolental fibroplasias, asphyxia, meconium aspiration, neurodegenerative disorders, chromosomal abnormalities, drug and alcohol abuse by the mother, maternal diabetes, multiple births, congenital infections, low birth weight (<1500g), bacterial meningitis, bilirubin > 10 umol/L, family history of hearing loss, craniofacial anomalies and ototoxic drug use > 6 days, are selected as impact factors.

International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:7, 2012

TABLE I
FACTORS INFLUENCING NEWBORN HEARING LOSS

Factors
length of stay in the intensive care unit-ICU>5 bays
respiratory distress syndrome
retrolental fibroplasias
asphyxia
meconium aspiration
neurodegenerative disorders
chromosomal abnormalities
drug and alcohol abuse by the mother
maternal diabetes
multiple births
congenital infections
low birth weight < 1500g
bacterial meningitis
bilirubin > 10 umol/L
family history of hearing loss
craniofacial anomalies
ototoxic drug use > 6 days

D. Data Analysis Software

This study selected and used the mining tool, Clementine V13.0 software, is developed by SPSS Company. Clementine was originally released as a reusable data mining working platform in 1994 by a British company, Integral Solutions Lts. (ISL). It could conveniently use complicated data mining algorithms and the required support functions, such as data access, pre-processing, graphs and reports, etc. Clementine has comprehensive graphs and models that provide classifications including SVM, logistic regressions, etc. Its expansibility is higher than the general software, so this study mainly adopted SVM and logistic regression to compare and conduct experimental analysis.

IV. RESULTS

A. Model Analysis Results

Model I—Support Vector Machine Computing Model

The first step is to integrate 563 pieces of patient data with 17 possible factors that may cause newborn hearing loss. Upon randomization, use 450 pieces of the training data and 113 pieces of the test data and input them to the model system. When the research variables and data are selected, this study firstly conducted SVM on the training sample data set to acquire the most appropriate parameters for the model. The results are as shown in Table II. Next, validate the accuracy of the constructed model based on the testing sample data set; record accuracy of each training and test data. To avoid overlaps on the dimension fittings on the test and K-fold training data of the parameters, at this stage, all the information will be randomized again for higher accuracy of the experiment. Then the optimal accuracy based on the results will be selected as the evaluation standard in this study to evaluate the best predictive model of SVM. Through SVM and K-fold Cross Validation Method, this study found accuracies of the 10 cluster groups are as listed in Table III. The average training accuracy is 96.409% and the average test accuracy is 96.09%. This study, based on the best accuracy by experiment validations, construct accurate clinical diagnostic models.

TABLE II
MODEL ONE PARAMETER SETTING

Parameter	Value
Computing Mode	RBF
Regression Accuracy	0.05
Gamma	0.2647
Normalized Parameter	2
Accuracy	94.69%
Error	5.31%

TABLE III
MODEL I-- K-FOLD CROSS VALIDATION ACCURACY

Cluster Group (K=10)	Training Accuracy (%)	Test Accuracy (%)
1	96.44%	96.49%
2	96.64%	94.74%
3	96.25%	98.25%
4	96.65%	94.64%
5	96.25%	98.21%
6	96.65%	91.07%
7	96.06%	100%
8	96.06%	96.43%
9	96.84%	92.86%
10	96.25%	98.21%
Average	96.409%	96.09%

Model II – Logistic Regression Combined with Support Vector Machine (SVM)

The first step is to randomly distribute the order of 563 data. Input 17 items of variables into logistic regression to compute and select the significant variables (a<0.5). According to its computing method acquired three kinds of variables: the forward method, the input method, and the backward method respectively. After conducting logistic regression analysis organized 6 and 3 items of significant variables respectively (see Table IV); then use these variable to conduct SVM computing. Divide 563 patient data in 450 training data and 113 test data for the test model parameters (see Table V). Among all, the prediction performance of parameters in the backward method is the best. Finally, though SVM computing and K-fold Cross Validation Method, the accuracy results of 10 cluster groups are as illustrated in TableVI. The average training accuracy is 96.565%, and the average test accuracy is 96.43%. This study based on experiments to validate the best accuracy in order to establish the accurate clinical diagnostic model.

B. Comparisons of the Experiment Models

Through the above two sets of experiment models obtained two sets of different data. Using logistic regression (the backward method) with SVM computing model presents the highest accuracy at 96.43%; the model I not going through variable filtering by logistic regression shows the accuracy at 96.09%. Compared to the best research outcomes in the past, Model II presents 3.18% higher accuracy in prediction ability.

TABLE IV
MODEL II SIGNIFICANT FACTORS IN LOGISTIC REGRESSION

Forward Method	Sig.	Input Method	Sig.	Backward Method	Sig.
Respiratory Distress	0	Respiratory Distress	0	Respiratory Distress	0
Meconium Aspiration	0.0891	Congenital Infections	0.004	Meconium Aspiration	0.0891
Congenital Infections	0.007	Bilirulin > 10 umol/L	0	Length of stay in the intensive care unit (ICU) > 5 days	0.005
Bacterial Meningitis	0.092	-	-	Congenital Infections	0.002
Bilirubin > 10 umol/L	0	-	-	Bacterial Meningiis	0.092
Ototoxic Drug Use > 6 days	0.01	-	-	Bilirubin > 10 umol/L	0

TABLE V
MODEL II PARAMETER SETTING

Parameter	Forward Method	Input Method	Backward Method
Computing Mode	RBF	RBF	RBF
Regression Accuracy	0.05	0.05	0.05
Gamma	0.75	1.5	0.75
Normalized Parameter	1	1	1
Accuracy	96.64%	94.74%	96.65%
Error	4.36%	5.26%	3.35%

TABLE VI
MODEL II K-FOLD CROSS VALIDATION METHOD ACCURACY

Cluster Group (K=10)	Training Accuracy (%)	Test Accuracy (%)
1	96.84%	94.74%
2	96.84%	94.74%
3	96.44%	94.74%
4	96.65%	98.25%
5	96.65%	96.43%
6	96.45%	94.64%
7	96.45%	94.64%
8	96.45%	94.64%
9	96.25%	98.21%
10	96.63%	98.21%
Average	96.565%	96.43%

TABLE VII
COMPARISONS OF MODEL ANALYSES

	Model I	Model II	Past Studies
Mining Mode	SVM	Logistic Regression with SVM	Decision Tree
Accuracy	96.09%	96.43%	93.25%

V. CONCLUSIONS

This study used newborn hearing loss as its topic, and its research subject being the real information provided by the case hospital, through the results of logistic regression and SVM of data mining, it is found that accuracies of all experiment models designed in this study are above 95%. Among all models, Model II, logistic regression using the backward method combined with SVM computing model has the highest accuracy at 96.43%, Model I at 96.09%; both has higher prediction accuracy than the past at 93.25%, indicating the function that filters variables in logistic regression has critical influence on the analysis of newborn hearing loss using SVM. Based on the two experiment models and past studies for verification, the weights of patients' clinical physiological conditions in both respiratory distress syndrome and bilirublin > 10umol/L are extremely high, making these two important indicators in judging newborn hearing loss. Referring to the factors discussed using artificial intelligence as one referential indicator in counseling and diagnosing newborn hearing loss, physicians can be provided with important information in clinical settings.

This study collected a total of 563 pieces of data provided by the case hospital and 17 variables for the research. However, it is still not enough in data collection used as statistical samples. Applications of the research results might have limitations due to the sources and characteristics of the number of samples collected. In the future, increasing number of samples by expanding regions of samples should be considered; more in-depth observations are needed to increase class variable items; alternatively, more research methods such as particle swarm optimization (PSO), case-based reasoning (CBR), and etc. should be added to make up the insufficiencies of logistic regression and SVM to improve accuracy of the predictive model.

REFERENCES

[1] Northern, J. L., & Hayes, D. H. (1994), "Universal screening for infant hearing impairment: necessary, beneficial and justifiable", Audiology Today, vol. 6, pp.10-13.
[2] Carney, A. E.,& Moeller, M. P. (1998), "Treatment efficacy: Hearing loss in children", Journal of Speech, Language, and Hearing Research, vol. 41, pp. 61-84.
[3] Northern, J. L., & Downs, M. P. (2002), Hearing in children (5th ed). Lippincott Williams & Wilkins., Philadelphia.
[4] Stilianos E. and Anestis Psifidis. and C. Joseph Chang. and Charles M. Stiern berg.(1997). "Risk Factors Associated With Hearing Loss in Neonates", American Journal of Otolaryngology, vol.18. pp.90-93
[5] Stilianos E. and Ioannis Skoulas. and Diane Phillips. and C. Y. Joseph Chang.(2002). "Risk Factors for Hearing Loss in Neonates: A Prospective Study," American Journal of Otolaryngology, vol.23. pp.133-137.
[6] Yoshinaga-Itano, C. (1999), "Benefits of early intervention for children with hearing Loss", Otolaryngologic Clinics of North America, vol. 32, pp. 1089-1102.
[7] Down, M.P., & Yoshinaga-Itano, C. (1999). "The efficacy of early identification and intervention for children with hearing impairment", Pediatric Clinics of North America, vol. 46(1), pp. 79-87.
[8] Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998) "Language of early- and later-identified children with hearing loss", Pediatrics, vol. 102(5), pp. 1161-1171
[9] Ross, M. (1990). "Implications of delay in detection and management of deafness", Volta Review, vol. 92, pp. 69-79.
[10] Kral, Hartmann, Tillein, Held & Klinke.(2002). Hearing after congenital deafness: central auditory plasticity and sensory deprivation, Cerebral cortex, USA.
[11] Rosemary, C. (1998). "Further support for the benefits of early identification", Volta Review, vol. 100(5), pp. 32-53.