Neural Networks for Recognizing Patterns in Cardiotocograms

Claudia Ulbricht, Georg Dorffner
Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A–1010 Vienna, Austria
and Department of Medical Cybernetics and Artificial Intelligence
at the University of Vienna
Freyung 6, A–1010 Vienna, Austria
claudia@ai.univie.ac.at, georg@ai.univie.ac.at

Andreas Lee
University of Vienna, AKH
Department of Obstetrics and Gynecology
Division of Prenatal Diagnosis and Therapy
Währinger Gürtel 18–20, A–1090 Vienna, Austria
Andreas.Lee@univie.ac.at

Abstract

The cardiotocogram (CTG) is commonly used for routine fetal monitoring in the delivery room. A major problem is that the interpretation of the CTG trace requires experienced specialists. In order to avoid long gaps between the detection of a suspicious pattern and the intervention, the CTG has to be checked in short intervals. An automated monitoring system at the obstetric site can reduce such delays. Therefore, an alarm system immediately reporting suspicious events has been built. The focus of our study was put on the question whether AI techniques such as neural networks are suited to the task of recognizing patterns in the CTG trace. In a comparative study, their performance was evaluated against that of conventional methods. The neural networks turned out to provide significantly better results than the tested conventional methods.

Keywords:

Neural Networks, Alarm System, Medicine, Gynecology, Obstetrics, Monitoring, Cardiotocography

1 Introduction

The use of the cardiotocogram (CTG) for routine fetal monitoring before birth, revealing important information concerning the health of the fetus, is wide-spread.

Several recent studies have shown that CTG monitoring reduces fetal mortality, but in contrast to physicians' expectation, it does not reduce fetal morbidity. Another problem is that this method requires continuous checking of the paper output produced by the CTG device. The interval between the detection of suspicious patterns in the CTG stream and the obstetric intervention is often too long. Any delay between the incidence of fetal distress and the required action increases the risk of a poor outcome. An automated tool immediately reporting suspicious situations helps in reducing such delays. As suggested in [Devoe, 1996], computerization can lead to more appropriate actions, because an automated analysis is objective and reproducible. Since such systems are well tested, their application in obstetric practice is reasonable. When designed appropriately, they can support the team during critical phases. As argued in [Greene, 1996], the goal is not replacing experts, but providing an "interactive decision support tool."

High inter- and intraobserver variability of manual CTG interpretation is another serious problem suggesting the use of an automated analysis tool [Keith and Greene, 1994]. However, most approaches to utilizing alarm systems fail, because the analysis produced by these systems does not correspond with that of the people working at the specific site. Since fixed practicable standards are missing, a flexible tool, which can be adapted to the specific environment, is required. The goal of our study is to investigate whether Artificial Intelligence methods (AI methods), such as neural networks (see, for instance, [Hertz *et al.*, 1991]), are suited to taking over some tasks within an automated tool for monitoring the CTG. In our experiments, a neural network was used for detecting so-called decelerations. These are patterns in the CTG, which are relevant for checking the situation of the fetus. Neural networks have already been applied to various tasks in the field of fetal monitoring, but they differ from our application, because they specialize on other subtasks. Experiments are reported, for instance, in [Kol *et al.*, 1994], [Devoe *et al.*, 1995], and [Ergün *et al.*, 1995]. All the other tasks of the automated monitoring tool were performed with conventional methods.

2 Data Sets

An important difference between our system and most existing systems is the way data is recorded. In the system 8000, the values are averaged over 1/16-minute intervals [Dawes *et al.*, 1991], while in the system presented in [Chung *et al.*, 1995], the values are sampled at intervals of ten seconds. In our system, each single heartbeat interval is stored. Our type of sampling is referred to as "irregular interval sampling." It stands in contrast to "fixed interval sampling" [Chung *et al.*, 1995]. A number of representative data sets were collected for the comparative study.

In our setting, the CTG monitoring device was connected to a PC, which stored the files. These files contain the intervals between subsequent fetal heartbeats (FHB values) in msec. For each such value the uterine contraction (UC value) in kPa was stored. The advantage of storing each single heartbeat is that various variability measures can be determined, which are relevant for recognizing patterns in the CTG trace. Most other systems work with heart rate values which are already pre-processed and smoothed. This way, the influence of single outliers is reduced and the resulting smoothed curve can be analyzed more easily. However, information hidden in the order of subsequent heartbeats is lost and can thus not be exploited. The difference between these two types of measuring is depicted in Fig. 1. The heart rate is close to the mirrored picture of the fetal heartbeat intervals, but due to sampling the curve is squeezed in the upper part, and stretched in the lower part.

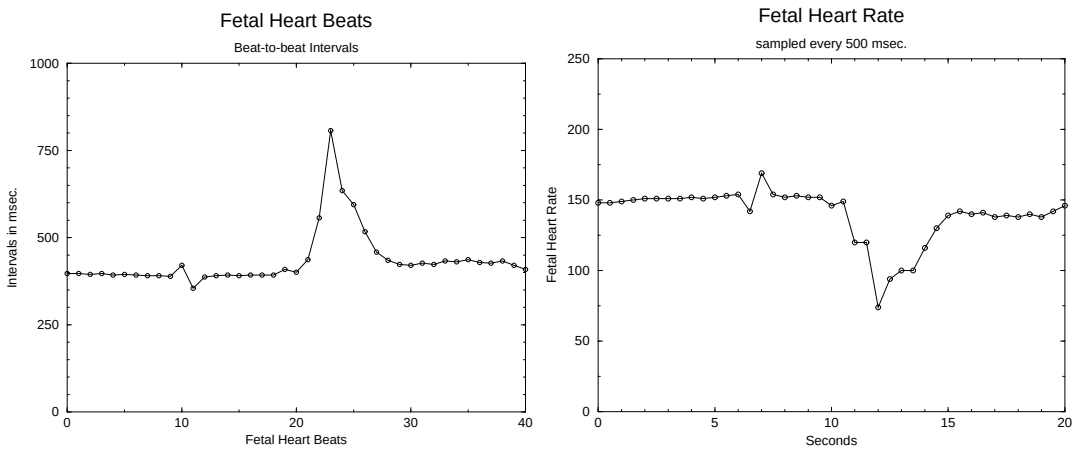

Figure 1: *Comparison of two types of data processing*

The collected CTG traces were inspected and marked by a single specialist. A program developed specifically for this task supported this work. Various points where an alarm should be given were marked as such. All the other steps of the CTG analysis were automated, so that the output of the CTG monitoring device can directly be forwarded to our system. This means that our system can be used on-line.

3 The Alarm System

In the on-line mode, each value measured by the monitoring device is forwarded to the alarm system, which checks a number of conditions. If any of these conditions is met, an alarm is produced. Theoretically, these modules can work in parallel.

In this type of system, though, the conditions are checked in a sequential order. In Fig. 2, an overview of the structure of the alarm system is given. The CTG signal is the only input. The first step includes the detection of artefacts, which are either removed or smoothed out. Then, relevant values, such as the new baseline value and the variability, are determined. They are available to the modules which check whether an alarm should be given. Finally, all the triggered alarms are reported without delay. The deceleration detection module is highlighted, because this is the module which is focused on in the described study. While the rest of the system remains unchanged, different algorithms for deceleration recognition can be plugged in and tested.

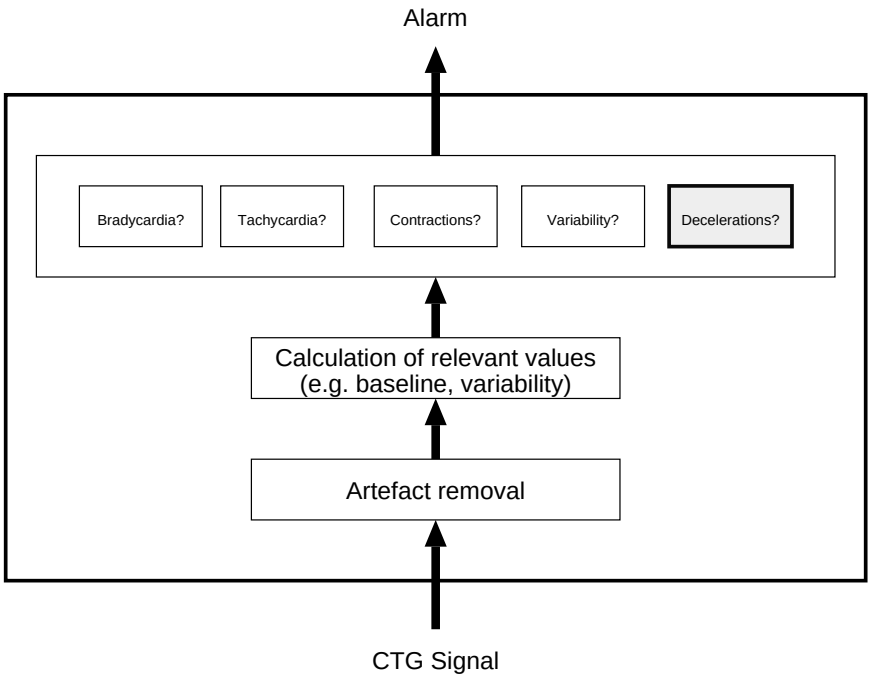

Figure 2: *The alarm system*

Artefacts: Each incoming value is checked for validity. If an FHB value lies outside the range of [250,1200] msec, it is regarded as an artefact and marked as such. This way, artefacts can be excluded from further calculations. As soon as a valid value is obtained, the preceding artefacts are replaced by interpolated values. The UC signal is also treated this way. The values have to lie in the range of [0,150] kPa. Moreover, outliers caused by single faulty FHB measurements are smoothed out. If a sudden increase larger than 35 bpm (beats per minute) is followed by a decrease of more than 70 bpm, and another increase of more than 35 bpm, all the involved points are replaced by averaged ones, so that they form a smooth line. Analogously, the signal is also checked for similar patterns

which start with a decrease rather than an increase. For these calculations, the FHB values are converted to FHR (fetal heart rate) values by determining the corresponding heart rate

$$r = \frac{60000}{b} \tag{1}$$

where r is the fetal heart rate in beats per minute, and b the length of the distance between two heartbeats in msec. This conversion allows the application of the FHR measures which are traditionally used for CTG traces.

Contractions: In the current system, a contraction is assumed when the UC value lies above 60 kPa. The time span between contractions is measured in order to detect frequent contractions. Intervals between 10 and 100 seconds trigger an alarm. Shorter ones are regarded as single contractions, disturbed by unfiltered artefacts as they are common during contractions.

Calculating the Baseline: Since the intervals between the fetal heartbeats are given, the baseline is also calculated in milliseconds. For further calculations, it is transformed to an FHR value. The quality of the baseline is very important for determining other measures which take the baseline into account. In order to obtain a reliable baseline, it is not recalculated if the preceding heartbeat interval is recognized as an artefact, as being of bad quality, as belonging to an acceleration, or as being part of a deceleration. In these cases, the old baseline value is taken over. In the beginning, the baseline is set to the first encountered interval to start in a reasonable range. This way, reasonable values are obtained right away.

The baseline was calculated in two different ways, and all the experiments were performed twice. For the first measure, the mean of the FHB values in the last 5 minutes is determined. Artefacts, values of bad quality, accelerations, and decelerations are excluded. In order to smooth the baseline, 1% of this new value is added to 99% of the preceding baseline value. The same algorithm was used for determining the baseline of the UC signal. The second type of baseline differs in that only a single preceding minute is taken into account. This is comparable to the way the baseline is determined in [Hamilton and Kimanani, 1994]. Again, artefacts, values of bad quality, accelerations, and decelerations are excluded. The baseline is not smoothed, though. But — like in [Hamilton and Kimanani, 1994] — the new value is only accepted if it is equal to at least 5% of the values in the minute interval.

Determining the Variability: The variability represents a measure for the size of the fluctuations. In [Oppenheimer and Lewinsky, 1994] it is argued, that the FHR variability is very important for determining the health status of the fetus before it is born. It is assumed that the beat–to–beat intervals contain

relevant information. Moreover, these intervals are heavily dependent on their order. Taking the mean is not appropriate.

Two different methods of calculating the variability were tested. The default method is based on the mean of the absolute differences between subsequent FHB values in the preceding 20 seconds interval. The second method uses the time period given in [Hamilton and Kimanani, 1994]. There, the mean of the differences between subsequent FHB values in the preceding 60 seconds interval is determined.

In both cases, the resulting measures of variability are more precise than those derived from smoothed FHR values. Since important information concerning the status of the fetus is assumed to be reflected in the variability of the heartbeat intervals, this approach is expected to lead to more reliable results.

Quality: In order to continuously monitor the quality of the incoming signal, the amount of artefacts in the last 5 minutes is determined. If it exceeds 30%, the signal is marked and an alarm is produced. The UC signal is checked analogously.

Bradycardia: In order to discover bradycardia, the FHB differences between adjacent intervals in the preceding 1-minute range which are smaller than 5 bpm are counted. Artefacts are excluded. If more than 5 valid values remain, and if at least 80% of them are small (i.e. < 5 bpm), a bradycardia alarm is given.

Tachycardia: If the baseline of the FHB signal reaches a fetal heart rate of 150 bpm in more than 80% of the valid cases within the last 120 seconds, and if at least 5 FHB values are valid, a tachycardia alarm is given.

Detecting Decelerations: Decelerations are of major relevance for the description of the CTG trace. They are characterized by a decrease of the fetal heart rate. An example is displayed in Fig. 1. Not only the *existence* of decelerations is substantial, but also the *number* of variable and non-variable decelerations per hour [Hamilton and Kimanani, 1994]. The automatic classification of decelerations is not trivial. Different criteria were used in the experiments:

- For the first method, the amount of low fetal heart rate values in the preceding 3 seconds is determined. A value is low if it lies at least 27 bpm below the baseline. In a preceding comparative study with an independent data set, these thresholds turned out to achieve the highest number of correct classifications. If the tested period contains too many low values, a deceleration alarm is triggered.

- The second method takes both shallow and deep decelerations into account. Like in the System 8000 [Dawes *et al.*, 1991], an alarm is generated if the

signal lies more than 10 bpm below the baseline for at least 60 seconds, or more than 20 bpm below the baseline for at least 30 seconds.

- According to the system described in [Chung *et al.*, 1995], the difference between the FHR signal and the baseline has to exceed 15 bpm for at least 20 seconds. A minimum of 5 valid measures is required for an alarm.
- The fourth method differs only in the thresholds. Like in the system described in [Kol *et al.*, 1994], the distance to the baseline has to exceed 10 bpm for a minimum of 15 seconds.

The practical applicability of these methods was tested in a comparative study, which is described in the following section.

4 Comparative Study

Our study mainly aimed at comparing different methods for recognizing decelerations. A number of conventional methods and neural networks were tested by letting them detect decelerations in CTG streams.

4.1 Conventional Methods

The collected data set was classified by different systems, which are the result of combining various techniques found in literature and newly developed algorithms. Four methods for detecting decelerations, and two algorithms for determining the baseline were tested. When they are combined, this results in eight different systems. The tested types of methods are listed in Table 1. The references indicate which systems inspired the respective approaches. They are all addressed in Sec. 3.

Symbol	Focus	Origin of Method
A	Deceleration	Kol et al. (altered thresholds)
B	Deceleration	Dawes et al.
C	Deceleration	Kol et al.
D	Deceleration	Chung et al.
a	Baseline	Hamilton & Kimanani (altered thresholds)
b	Baseline	Hamilton & Kimanani

Table 1: *Types of methods*

An overview of the criteria underlying the tested methods can be found in Table 2. The minimum width (in seconds) and depth (in beats per minute) are given. A deceleration is recognized as such if the width and the depth of the detected pattern exceed the given thresholds.

Method	Width	Depth
A	3 sec.	27 bpm.
B	60 sec.	10 bpm.
	or 30 sec.	20 bpm.
C	15 sec.	10 bpm.
D	20 sec.	15 bpm.

Table 2: *Description of methods*

Results: The methods are referred to as methods 'A' through 'D' for different types of deceleration recognition, and 'a' and 'b' for different types of baseline calculation. The criteria found in literature (methods 'B,' 'C,' 'D,' and 'b') turned out not to be well suited to our data sets. Only a very small percentage of the decelerations were recognized. This demonstrates the weakness of fixed criteria. They may be useful in a specific environment, but not in general. In our environment, they turned out to be not applicable. Therefore, the thresholds were adjusted until the number of correctly classified patterns reached a maximum. Method 'A' is the result of a set of experiments with an independent data set. There, the thresholds used by method 'A' turned out to lead to the best performance. Similarly, the baseline algorithm referred to as method 'a' provided better results than the original algorithm (method 'b').

The criteria used in the System 8000, which is described in [Dawes *et al.*, 1991], and in method 'B' are more sophisticated than the others in that two types of patterns are distinguished. Decelerations can be short and deep, or long and shallow. However, the results are poor in spite of these more elaborated decision rules. Much better results were obtained when employing the adjusted criteria of method 'A.' We have labeled this method a "conventional method," because this group stands in contrast to the group of neural network methods, but it should be pointed out that the applied criteria are already part of the outcome of our research. The fact that our criteria yield better results can be interpreted in different ways. The standard criteria may not be optimal, or the criteria used at our site deviate from the standard. The high inter-observer variability regarding the interpretation of CTG traces is well known [Keith and Greene, 1994] and poses a major problem for the automation of the CTG inspection.

An overview of the results is provided in Table 3. The total number of correctly classified decelerations is given in the rightmost column. Moreover, the number of correct positive, false negative, false positive, and correct negative classifications is provided. It can be seen that in all cases the number of false negatives is rather high. This may be due to the use of the original FHB values instead of smoothed FHR values. Another reason is that the applied criteria are specific to the clinic and the physician. Standard criteria are not sufficient, because the variance regarding the classification of decelerations is quite high.

	Method		Results				Total
	Dec.	Basel.	C. pos.	F. neg.	F. pos.	C. neg.	Correct
1	A	a	50%	50%	0%	100%	75%
2	B	a	4%	96%	0%	100%	52%
3	C	a	34%	66%	6%	94%	64%
4	D	a	14%	86%	0%	100%	57%
5	A	b	52%	48%	2%	98%	75%
6	B	b	2%	98%	0%	100%	51%
7	C	b	36%	64%	8%	92%	64%
8	D	b	14%	86%	2%	98%	56%

Table 3: *Overview of the results obtained with conventional methods*

Since neural networks learn from examples, they are more flexible and adapt themselves to the given environment. Due to this strength, they were expected to yield better results. The experiments with neural networks are described in the following section.

4.2 Neural Networks

The tested neural networks used here for detecting decelerations are all based on multi-layer perceptrons. Among them are also simple recurrent networks with hidden layer feedback. Such networks have already been used in a series of experiments, which is documented in [Ulbricht *et al.*, 1995]. They are comparable to the network presented in [Elman, 1990]. The basic network model has fifteen input units, three hidden units, and two output units. The two output units are trained to indicate decelerations by their activations. The network architecture is shown in Fig. 3. The dashed arrows denote connections for simply copying unit activations. The full arrows represent connections between all the units of the connected layers, which carry the weights of the neural network. The layers are numbered L1 through L4. This indicates the order in which the layers are

updated. The three hidden units are fed back to the state layer units, which can be regarded as a memory. Another network has additional self-recurrent feedback loops around the state layer units. They enable the network to deal even better with temporal aspects. The issues concerning the handling of temporal patterns with neural networks are discussed in detail in [Ulbricht *et al.*, 1992]. The other tested networks are variants of these networks with different numbers of units and weights.

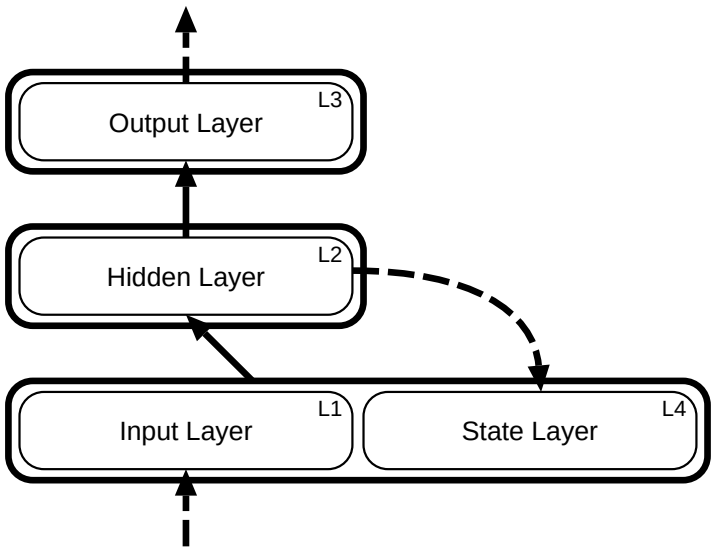

Figure 3: *A recurrent neural network with hidden layer feedback*

Cross-Validation: The neural networks were trained with the examples in a training set. An independent inspection set was only used for determining when to stop the training process. This avoids that the network adjusts too well to the examples in the training set. When over-fitting the training data, the generalizing capability of the network is not optimal, and thus its performance on new examples is not as good as it can be. The performance on an independent test set is then taken as a measure of the quality of the neural network. Due to the lack of a large set of examples, 80% of the examples were taken for training, 10% for deciding when to stop the training process, and 10% for testing. This is visualized in Fig. 4. Then, these sets were rotated ten times and each time a new network was trained. In all the tables and figures, the mean value of performance is reported. This ten-fold cross-validation provides a much more reliable measure of performance on independent test sets than a test with a single test set.

In order to get a well-performing network, a set of so-called "near-misses" was collected. They share some features with decelerations and can thus easily

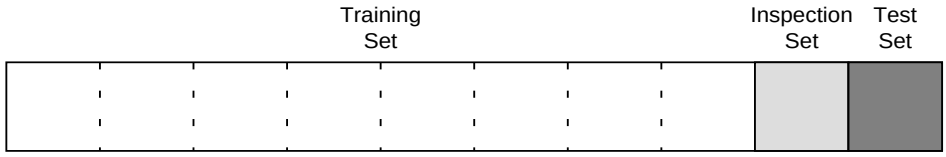

Figure 4: *Ten-fold cross-validation*

be misclassified as such. Together with examples of "normal" segments, they form the class called "no decelerations." This way, the neural network can be trained to distinguish two classes: "decelerations" and "no decelerations."

Neural Network Training: The inputs to the neural network are taken from the values calculated in the pre-processing phase of the alarm system. The points for training were selected randomly during training. Positive and negative examples were alternated. In order to adjust the state layer, the network was updated ten times before it was actually trained. For testing, complete time series were used. This is comparable to a true original incoming data stream as it can be found in a real world environment.

The strategy of evaluating the response of the alarm system is shown in Fig. 5. If a single deceleration alarm was given where no deceleration was present, the pattern was counted as a wrongly classified pattern (false positive). Accordingly, a single recognition of a deceleration in the marked interval sufficed to count the pattern as correctly classified (correct positive).

Results: The ten-fold cross-validation was performed for different types of neural networks. The results obtained with the neural networks can be found in Table 4. The mean performance on the independent test set, i.e. the percentage of correctly classified examples, lies between 71 and 76%. The minimum of correctly classified patterns in a single subset was 50%, the maximum 100%. The mean amount of correct positives lies between 72 and 90%, and that of false positives between 20 and 40%. The influence of the chosen algorithm for calculating the baseline is rather small. Therefore, the results obtained with method 'b' are close to those of method 'a.'

The results of all methods are visualized in Fig. 6. It shows the false positive and the correct positive rate. The optimal performance is that of a system with a correct positive rate of 1.0, and a false positive rate of 0.0. Such a system recognizes all the decelerations and produces no false alarms. Methods 'B' and 'D' do not invoke any false alarms, but they turn out not to be well suited to the given task, because only 2 to 14% of the decelerations are found. Method 'C' recognizes 34 to 36% of the decelerations, and produces a few false alarms. The

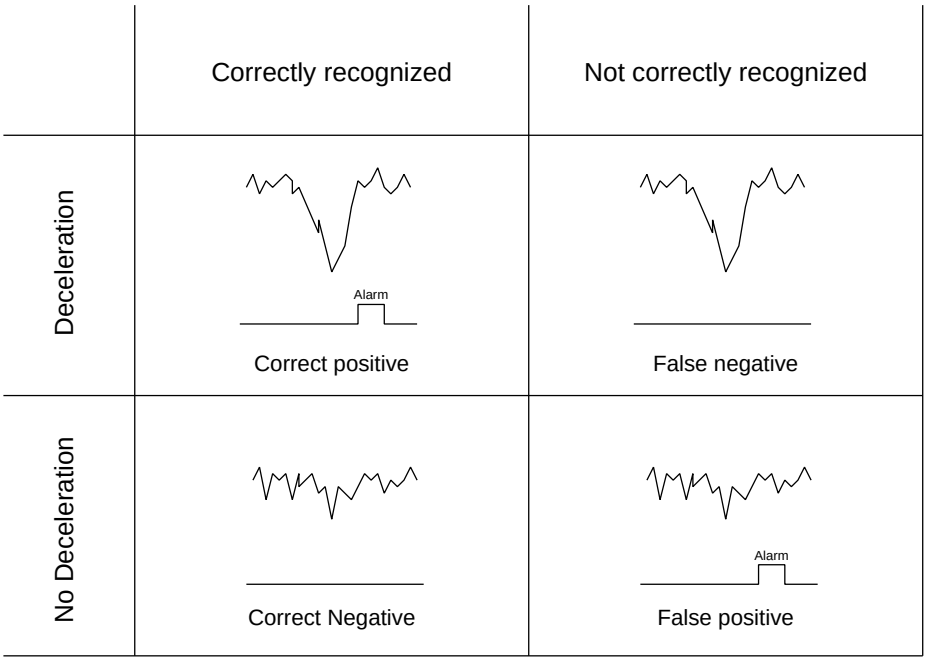

Figure 5: *The evaluation strategy*

	Tested Method	Results				Total Correct
		C. pos.	F. neg.	F. pos.	C. neg.	
9	Net1	72%	28%	20%	80%	76%
10	Net2	84%	16%	40%	60%	72%
11	Net3	90%	10%	40%	60%	75%
12	Net4	82%	18%	36%	64%	73%
13	Net5	88%	12%	36%	64%	76%
14	Net6	82%	18%	38%	62%	72%
15	Net7	78%	22%	32%	68%	73%
16	Net8	86%	14%	44%	56%	71%
17	Net9	88%	12%	36%	64%	76%
18	Net10	82%	18%	32%	68%	75%

Table 4: *Overview of the results obtained with neural networks*

thresholds of the newly introduced method 'A,' which is obtained by adapting the thresholds, turn out to be better suited. All the negative and half of the positive examples are correctly classified.

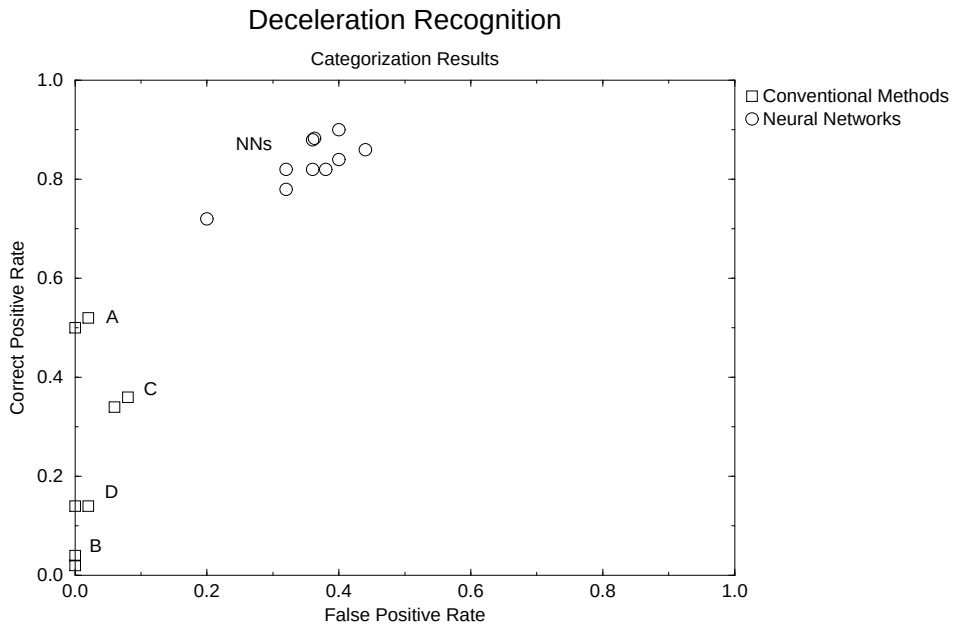

Figure 6: *Deceleration recognition with conventional methods and neural networks*

The points of the results obtained with the ten neural networks form a cluster which is clearly separated from that of the conventional methods. The t-Test (with a significance level of 0.05) reveals that even the network with the smallest number of correct positive results (Net1) is significantly better (concerning the correct positives) than the conventional methods 'B,' 'C,' and 'D.' Some of them (e.g.: Net 5, Net 8) can also be shown to be significantly better than method 'A.' Concerning the total number of correct classifications, the network with the smallest number of correct classifications (Net 8) is still significantly better than the conventional methods 'B,' 'C,' and 'D.'

4.3 Further Validation

In order to further compare the capabilities of conventional and neural network methods, the thresholds influencing the categorization boundaries were varied. The goal was to cover the whole range from a low false positive rate to a high correct positive rate. The resulting curves show how the rates can be tuned. They are referred to as ROC–curves (Receiver Operating Characteristics) [Centor and Keightley, 1989]. For testing the conventional methods, the thresholds of the

depth and the width were varied. Analogously, the thresholds of the output units in the neural network were tuned.

Results: In Fig. 7, the performance of the resulting system with adapted thresholds and a newly trained neural network is displayed. The curve of the neural network performance shows the mean of the results obtained with the ten-fold cross-validation. Both methods were tested with the same set of data sets. It can be seen that the conventional approach is better suited when the false positive rate should be low. Neural networks perform better when the correct positive rate is to be maximized.

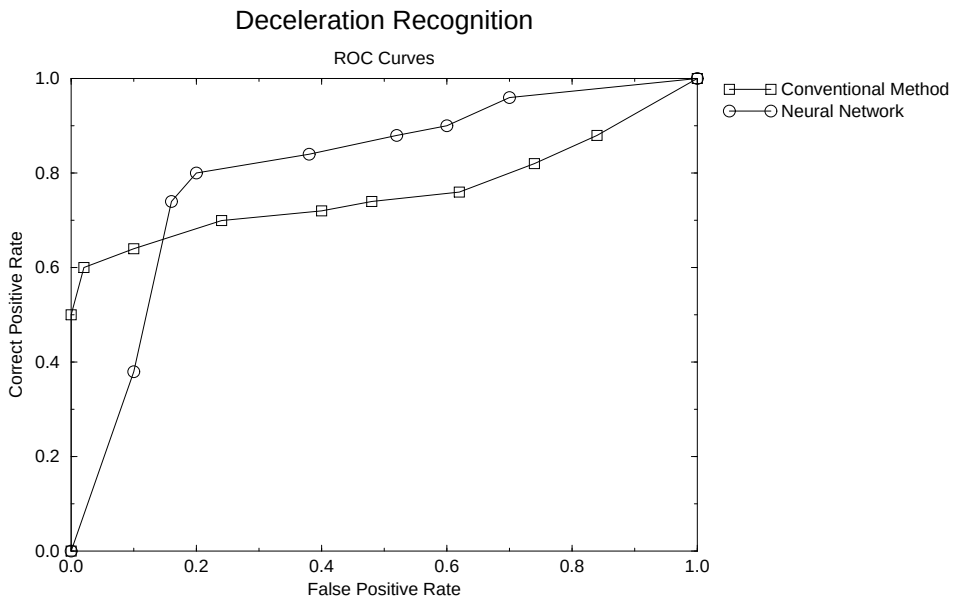

Figure 7: *ROC curves of a conventional method with variable thresholds and a neural network with variable decision boundaries*

For the t-Test, three systems with adapted thresholds and three systems with neural networks were selected. They all have a false positive rate in the range of [0.2,0.6]. The neural networks turned out to yield significantly better results concerning both correct positive and total correct classifications (for a significance level of 0.05). This shows that neural networks can indeed outperform the conventional methods. Thus, it makes sense to consider the application of neural networks to the detection of decelerations.

5 Discussion

A major strength of neural networks is their flexibility. They can be trained with arbitrary examples to categorize new patterns accordingly. As a result, they incorporate the decision rules used by the specialists rather than rough rules, which do not cover the whole task. The trained neural networks turned out to be better suited than predetermined rules.

On the other hand, physicians do not feel comfortable when relying on "black box" methods like neural networks. As argued in [Devoe, 1996], they miss the explanation capability of rules, which makes them trustworthy. Rules ensure that decelerations exceeding the given thresholds are definitely found. The mapping modeled by neural networks follows the statistical distribution of the data. Since existing rules work quite well, it may be more reasonable to employ neural networks only in addition to the conventional methods. However, this requires a measure of reliability. Whenever the reliability of the proposed decision is low, the network is consulted. The reliability measure could be dependent on the closeness of the parameters to the thresholds.

In comparison to most other methods, the main advantage of neural networks is that they can detect non-linear relations. In many cases, linear methods suffice, but in others, systems modelling non-linear mappings are much better suited. And then, neural networks yield better results. Also in our study, non-linear relations may be a reason for their superiority.

Neural networks have another strength. They can be employed for detecting relevant input features. For instance, when the results are worse after removing a single parameter from the inputs, this is an indication that this parameter may be important. In the presented study, the variability measures turned out to be relevant to the categorization of decelerations. When each single fetal heartbeat interval is recorded, the quality of the variability measures is higher than when the curve is smoothed. Various variability measures can also be used for refining existing rules. This way, neural networks can contribute to a better understanding of the issue of recognizing decelerations, and thus to the development of better automated tools.

As stated before, an advantage of employing neural networks is that they adapt themselves to the requirements of the specific environment. They can deal with the properties of the given monitoring device. They can also be adjusted to the standards of the clinic where the CTG is monitored. However, it is not clear whether the effort required for training a neural network to do this job is justified. Collecting instances, experimenting with different sets of input parameters and various neural network architectures is quite time-consuming. And the adaptation of the thresholds in conventional systems also resulted in well-performing systems.

The experiments showed that the influence of the chosen network architecture is rather low. All the tested networks yielded similar results. However, the

selected set of input parameters has a large effect on the performance. The currently used input parameters stem from the alarm system and are not tailored towards the task. Parameters which describe features of decelerations may be better suited. For instance, the width, the depth, and the steepness can be calculated in advance and supplied to the network. If the description of the investigated patterns resembles that used by the human investigators, it may be easier to obtain similar categorization criteria and similar performance. This could be the focus of future experiments.

Finally, it has to be pointed out that the number of training instances was rather small. More experiments with larger data sets are necessary to further validate the current results.

6 Conclusion

Due to missing practicable standards and to the high inter-observer variability, the interpretation of the CTG remains a difficult task. The overall goal is to avoid unnecessary delays by using automated alarm systems to support the decision process. Therefore, a system is needed which is actually well suited to supporting the work at the obstetric site.

The question posed in the beginning—whether neural networks are applicable to the task of CTG monitoring—can be answered positively. It has been shown that neural networks can be used for detecting patterns in the CTG trace. They yielded significantly better results than conventional methods. Therefore it can be expected that neural networks will also be applicable to numerous other subtasks within CTG monitoring.

Acknowledgements

This research is funded by the "Medizinisch–Wissenschaftlicher Fonds des Bürgermeisters der Bundeshauptstadt Wien."

References

[Centor and Keightley, 1989] R.M. Centor and G.E. Keightley. Receiver Operating Characteristic (ROC) Curve Area Analysis Using the ROC ANALYZER. In L.C. Kingsland, editor, *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care (SCAMC-89)*, pages 222–226, Washington D.C., 1989. IEEE Computer Society Press.

[Chung *et al.*, 1995] T.K.H. Chung, M.P. Mohajer, Z.J. Yang, A.M.Z. Chang, and D.S. Sahota. The Prediction of Fetal Acidosis at Birth by Computerised

Analysis of Intrapartum Cardiotocography. *British Journal of Obstetrics and Gynaecology*, 102:454–460, 1995.

[Dawes *et al.*, 1991] G.S. Dawes, M. Moulden, and C.W.G. Redman. System 8000: Computerized Antenatal FHR Analysis. *J. Perinat. Med.*, 1991.

[Devoe *et al.*, 1995] L.D. Devoe, E. Carlton, and P. Prescott. Neural network prediction of nonstress test results: How often should we perform nonstress tests? *Am J Obstet Gynecol*, 173:1128–1131, 1995.

[Devoe, 1996] L.D. Devoe. Computerized fetal heart rate analysis and neural networks in antepartum fetal surveillance. *Current Opinion in Obstetrics and Gynecology*, 8:119–122, 1996.

[Elman, 1990] J.L. Elman. Finding Structure in Time. *Cognitive Science*, 14:179–211, 1990.

[Ergün *et al.*, 1995] G. Ergün, M.A. Erkmen, M.S. Beksac, K. Izzetoglu, and U. Karaks. Evaluation of Fetal Heart Rate Tracings Based on Feature Extraction by Wavelet Transformation. *Gynecology Obstetrics and Reproductive Medicine (GORM)*, 1(1):5–7, 1995.

[Greene, 1996] K.R. Greene. Intelligent fetal heart rate computer systems in intrapartum surveillance. *Current Opinion in Obstetrics and Gynecology*, 8:123–127, 1996.

[Hamilton and Kimanani, 1994] E. Hamilton and E.K. Kimanani. Intrapartum Prediction of Fetal Status. *Clinical Obstetrics and Gynaecology, International Practice and Research, Analysis of Complex Data in Obstetrics, M.S. Rogers, A.M.Z. Chang, Bailliere Tindall, Volume 8/Number 3*, 1994.

[Hertz *et al.*, 1991] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, 1991.

[Keith and Greene, 1994] R.D.F. Keith and K.R. Greene. Development, Evaluation and Validation of an Intelligent System for the Management of Labout. *Clinical Obstetrics and Gynaecology, International Practice and Research, Analysis of Complex Data in Obstetrics, M.S. Rogers, A.M.Z. Chang, Bailliere Tindall, Volume 8/Number 3*, 1994.

[Kol *et al.*, 1994] S. Kol, I. Thaler, N. Paz, and O. Shmueli. Interpretation of Nonstress Tests by an Artificial Neural Network. Technical report, Department of Obstetrics and Gynecology, Rambam Medical Center, and the Faculties of Medicine and Computer Sciences, Technion, Israel Institute of Technology, 1994.

[Oppenheimer and Lewinsky, 1994] L.W. Oppenheimer and R.M. Lewinsky. Power Spectral Analysis of Fetal Heart Rate. *Clinical Obstetrics and Gynaecology, International Practice and Research, Analysis of Complex Data in Obstetrics, M.S. Rogers, A.M.Z. Chang, Bailliere Tindall, Volume 8/Number 3*, 1994.

[Ulbricht *et al.*, 1992] C. Ulbricht, G. Dorffner, S. Canu, D. Guillemyn, G. Marijuán, J. Olarte, C. Rodriguez, and I. Martin. Mechanisms for Handling Sequences with Neural Networks. In C.H. Dagli et al., editors, *Intelligent Engineering Systems through Artificial Neural Networks, ANNIE'92*, volume 2, pages 273–278. ASME Press, New York, 1992.

[Ulbricht *et al.*, 1995] C. Ulbricht, G. Dorffner, and A. Lee. Computer Supported Evaluation of the CTG with Neural Networks. In A. Pedotti and P. Rabischong, editors, *Proceedings of the 3rd European Conference on Engineering and Medicine, European Society for Engineering and Medicine*, page 435, 1995.