Artificial neural network modelling of the results of tympanoplasty in chronic suppurative otitis media patients

Joanna Szaleniec, Maciej Wiatr, Maciej Szaleniec, Maciej Składzień, Jerzy Tomika, Krzysztof Olesa, Ryszard Tadeusiewicz

Department of Otolaryngology, Jagiellonian University Medical College, ul. Sniadeckich 2, 31-501 Krakow, Poland
Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, ul. Niezapominajek 8, 30-239 Krakow, Poland
AGH University of Science and Technology, ul. Mickiewicza 30, 30-059 Krakow, Poland

Abstract

The application of computer modelling for medical purposes, although challenging, is a promising pathway for further development in the medical sciences. We present predictive neural and k-nearest neighbour (k-NN) models for hearing improvements after middle ear surgery for chronic otitis media. The studied data set comprised 150 patients characterised by the set of input variables: age, gender, preoperative audiometric results, ear pathology and details of the surgical procedure. The predicted (output) variable was the postoperative hearing threshold. The best neural models developed in this study achieved 84% correct predictions for the test data set while the k-NN model produced only 75.8% correct predictions.

1. Introduction

1.1. Artificial neural networks in the modelling of pathological processes

Computer modelling of pathological processes remains an underestimated tool in routine clinical practice. Nevertheless, the use of model-based procedures in technology, economics and science (e.g., physics, chemistry, astronomy and, recently, biology [1–4]) is widely appreciated. This diverse usage of models suggests that the application of computer modelling for medical purposes, although challenging, is a promising pathway for further development in the medical sciences. A medical model can aid treatment planning and optimisation, provide predictions of treatment results, facilitate decision-making processes or serve as a tool for medical education.

There is a fundamental difference between modelling in physics or technology and modelling in medicine. For most pathological processes, the quantitative relationship between causes and consequences (results) remain unclear. The optimal solution to this problem is the development of medical knowledge, which could lead to a better understanding of physiology and pathology, thus enabling progress from qualitative to quantitative descriptions of empirical observations. Perhaps in the future, medical research will provide mathematical formulas that describe all of the biological processes so that optimal treatment methods can be found by solving mathematical equations.

However, such detailed mathematical descriptions of most pathophysiological processes do not currently exist. In this situation, methods of artificial intelligence (AI) can be successfully applied to build models of biological processes without complete knowledge of the cause–effect relationships or the formulas that precisely describe these relationships [5]. One of the best AI methods for this purpose is neural networks technology [6]. In this paper, we show how this emerging technology can be used for a representative medical problem. The purpose of this paper is not only to present this particular problem and its solution by neural networks but also to describe step by step how to use neural networks for disease and treatment modelling, how to process clinical data for neural network learning and how to interpret output signals obtained from a neural network to answer important medical questions.

1.2. Representative medical problem

Chronic suppurative otitis media (CSOM) is the chronic inflammation of the middle ear. It is one of the most common causes of hearing impairment in the world. The disease presents with hearing loss and ear discharge through a tympanic perforation [7,8].
The treatment of choice for most cases of CSOM is surgery with the principal aim of eliminating the disease from the tympanic cavity and the mastoid process. The pathological changes found in the middle ear in CSOM include cholesteatomas (also called “pearl tumours”, which have the potential to destroy the adjacent bone), granulation tissue, adhesions and calcifications (tymanosclerosis). If untreated, these changes may lead to serious and sometimes life-threatening complications [9]. Common complications include facial nerve paralysis, labyrinthitis, lateral sinus thrombophlebitis, meningitis and brain or cerebellar abscesses [7,10,11].

Secondary goals of the ear surgery include hearing improvements and middle ear biology restoration. These goals are achieved through tympanic membrane and auditory ossicle reconstructions [12]. However, in many cases, the damage caused by the inflammatory process is so profound that achieving any hearing improvement is impossible [13]. Additionally, the removal of advanced lesions from the tympanic cavity or complications of surgery may lead to further hearing impairment instead of hearing restoration [14].

Definitive hearing results cannot be evaluated immediately after surgery. Because the healing process takes time, hearing improvement should be assessed after a follow-up period of at least several months [15]. However, patients desire to know whether the improvement should be expected. Moreover, if hearing function is significantly worse than anticipated for a particular patient, it may indicate potential complications or a failure of reconstruction that may require additional operations.

The aim of this paper is to present neural predictive models for hearing improvement after middle ear surgery for chronic suppurative otitis media. We discuss the methods of model optimisation and pitfalls that should be avoided as well as the potential goals of further research in the field.

1.3. General assumptions

There are many prognostic factors that affect hearing improvement after tympanoplasty. The prognosis depends on the pre-operative hearing levels, auditory tube function, the presence of ear discharge (infection), pathological changes found in the ear, the extent of the tympanic membrane and auditory ossicle destruction, the surgical technique, and the surgeon’s experience [16–18]. These factors have been statistically correlated with prognosis in large groups of patients. However, to our knowledge, no neural prognostic model has been developed to predict hearing improvement in individual patients.

The most frequently used hearing test is a pure-tone audiometry, whose results can be graphically recorded as an audiogram (Fig. 1). Hearing thresholds are measured for two factors: (i) air conduction (i.e., the conduction of sound waves to the inner ear through the outer ear and middle ear) and (ii) bone conduction (i.e., the conduction of sound waves to the inner ear through the bones of the skull). The air conduction threshold (ACT) depends on the function of the outer, middle, and inner ear, and the bone conduction threshold (BCT) principally provides a measure of the inner ear function that is, to some extent, distinct from the middle and outer ear function. The difference between the thresholds for air and bone conductions (“air-bone gap”—ABG) is therefore informative of the middle and outer ear function. Most authors who have analysed the effects of middle ear surgery on hearing report postoperative changes in the ABG. However, positive and negative changes in bone conduction after tympanoplasty have also been observed [19,20]. Therefore, we decided to report the air conduction threshold shift as a total measure of hearing improvement after tympanoplasty [15]. In our judgment, it is reasonable to refer to the air conduction threshold in the postoperative hearing prediction because (i) it describes the total effect of treatment on both the middle and inner ear, as it is the sum of the bone conduction threshold and the ABG and (ii) it reflects hearing in real-life situations and, as such, is more comprehensible for the patients than the ABG.

2. Materials and methods

2.1. Study population

The studied data set comprised 150 patients who underwent surgery for chronic suppurative otitis media between 2004 and 2007 in the Department of Otolaryngology, Jagiellonian University, Krakow. The data set included 84 women and 66 men aged 9–80 years (mean age 41 years).

Each patient provided informed consent for treatment and participation in the study. The analysis involved patients who underwent tympanoplasty surgery for the first time. Patients who previously underwent surgery on the same ear or who underwent other surgical procedures, as well as those who failed to complete a 12-month follow-up, were excluded from the study.

2.2. Analysed variables

Each patient was characterised by the following set of 21 independent variables (descriptors or input variables):

- age
- gender
- audiometric results (the value of preoperative bone conduction and preoperative ABG measurements; the mean value of three measurements at the speech frequencies of 500, 1000 and 2000 Hz)
- ear pathology (discharge, perforation or retraction pocket of the tympanic membrane, the presence of cholesteatomas, granulations, adhesions, tymanosclerosis, ossicular chain destruction or disconnection)
- surgical procedure descriptors (surgical approach, extent of surgery in the middle ear, type of material used for myringoplasty (tympanic membrane reconstruction) and ossiculoplasty (ossicular chain reconstruction))
Age, audiometric results and perforation size were expressed as continuous values and were normalised for use in the models. The remaining descriptors were discrete values and were coded with either a binary method (0 or 1) or a one-of-N method.

The predicted parameter was derived from the change in the air conduction threshold (ΔACT) resulting from surgery

\[
\Delta ACT = ACT_{\text{post}} - ACT_{\text{pre}}
\]

where \(ACT_{\text{pre}}\) and \(ACT_{\text{post}}\) are the air conduction thresholds measured before surgery and 12 months after surgery, respectively. ΔACT shows the postoperative hearing improvement (ΔACT < 0) or deterioration (ΔACT > 0). To simplify the prediction task, the ΔACT was binary encoded (1—hearing improvement, 0—no improvement) and was used as the dependent (output) variable. According to this classification, 111 patients in the study group exhibited hearing improvement, and 39 showed no improvement or suffered hearing deterioration.

Detailed descriptions of the variables can be found in Table 1.

To predict the output variable, two types of classification models were used: the feed-forward artificial neural networks and k-nearest neighbour models.

### Table 1

<table>
<thead>
<tr>
<th>Description of input variables and their encoding.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td><strong>Audiometric results</strong></td>
</tr>
<tr>
<td>Preoperative bone conduction</td>
</tr>
<tr>
<td>Preoperative air-bone gap</td>
</tr>
<tr>
<td><strong>Ear pathology</strong></td>
</tr>
<tr>
<td>Discharge</td>
</tr>
<tr>
<td>Incus destruction</td>
</tr>
<tr>
<td>Stapes destruction</td>
</tr>
<tr>
<td>Perforation</td>
</tr>
<tr>
<td><strong>Surgical procedure descriptors</strong></td>
</tr>
<tr>
<td>Surgical approach</td>
</tr>
<tr>
<td>Type of material used for myringoplasty (tympanic membrane reconstruction)</td>
</tr>
<tr>
<td>Cholesteatoma</td>
</tr>
<tr>
<td>Disconnection</td>
</tr>
<tr>
<td>Tympanoplasty</td>
</tr>
<tr>
<td>Adhesions</td>
</tr>
<tr>
<td>Granulation</td>
</tr>
<tr>
<td>Tympanosclerosis</td>
</tr>
<tr>
<td>Mastoidectomy</td>
</tr>
<tr>
<td>Ossiculoplasty (ossicular chain reconstruction)</td>
</tr>
<tr>
<td>Type of high columnella reconstruction</td>
</tr>
</tbody>
</table>

### 2.3. Artificial neural networks used in the experiments

All classification networks used in the experiments were created, trained and investigated using the Statistica® Neural Networks 7.1 and 9.0 software (www.statsoft.pl). All cases were divided into 3 subsets: learning, validation and testing. The partitioning into subsets was conducted after preprocessing based on k-mean clustering of the data set into 21 clusters. The clustering procedure considered all the variables collected in the database, including the postsurgical audiometric results. Analysis of the individual subgroups of cases after clustering revealed that each cluster comprised patients with similar preoperative and postoperative audiometric results, age and perforation size. The qualitative variables were roughly randomly distributed in the clusters.

The number of clusters was selected empirically taking into account two opposing premises: (i) the clusters should correctly represent each subgroup of patients (the smaller clusters the more detailed representation) and (ii) the minimal number of cases in each cluster has to allow a proportional selection of test and validation cases from every cluster (the minimal number of cases in the cluster depends on the assumed partitioning ratios). Therefore, the selected number of clusters into which the data are to be partitioned should provide a maximum number of clusters with at least 3–5 cases in each cluster. Although it is not typically possible to completely avoid classification of single cases into separate clusters (i.e., outlier cases), such situations can be avoided by decreasing the number of clusters. In our partitioning scheme, the average cluster population was 7.2, and we had only one cluster with a single case and two clusters with three cases. Small clusters must be included in the training or, if possible, test subsets. The cases from each cluster were partitioned separately into training, validation and test subsets in a 88:3:7:25 ratio on the basis of the distance from the cluster centre. This procedure ensured that the validation and test subsets were representative of the entire group.

The optimisation of network architecture is a crucial step in every neural network application. In the current research, this optimisation was conducted automatically with the Statistica 7.1 Intelligent Problem Solver (IPS). During the optimisation process, 1000 different network architectures were created and investigated. Our presumption was that only Multi-Layered Perceptron (MLP)-type networks will be taken into account. Other neural network architectures are possible; however, our previous experience showed that the three-layered MLPs are the most promising for the current problem. The IPS optimisation of the neural architecture was based on a minimisation of the network validation error and involved the selection of both the optimal input vector and the number of hidden neurons. The procedure used the uniform training method (i.e., 100 epochs of a back-propagation algorithm, 20 epochs of a conjunct gradient and 20 epochs of a conjunct gradient with momentum).

The best model selected by the IPS procedure, termed model 1, had the following architecture: MLP 17:30–8–1.1, which indicates that 17 input variables were encoded by 30 input neurons, that 8 hidden layer neurons were used for data processing and that 1 output neuron provided the prediction. Hyperbolic tangent and logistic activation functions were used for hidden and output neurons, respectively. The IPS procedure eliminated 3 input variables from the initial input vector which were highly correlated with the rest of parameters or occurred insignificant for the studied problem. The difference between the number of inputs and the number of input neurons originates from the 1-of-n representation of qualitative input data (see Table 1).

The model was manually optimised through the sensitivity analysis. As a result, the least important input parameter...
(the surgical approach) was removed, and the network (model 2, MLP 16:28-8-1:1) was manually retrained with 47 epochs of back propagation with momentum. Further reduction of the input vector by removing the next least important variable significantly deteriorated the model's quality. Furthermore, elimination of any strongly correlated variable also resulted in a decrease in the network's robustness. Finally, reduction of the number of hidden neurons did not improve the network's performance. In the last stage, the obtained model was subjected to further optimisation in the Neural Network module in Statistica 9.0. The optimisation varied the number of hidden neurons and activation functions without changing the input vector. The best model (model 3) was MLP 16:38-9-2:1 (the difference in the number of input and output neurons is due to a different coding system of binary categorical variables between Statistica 7.1 and 9.0). The network was trained with 17 epochs of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. The activation functions for hidden neurons and the output neuron were the hyperbolic tangent and softmax functions, respectively.

2.4. k-Nearest neighbour (k-NN) models

The k-NN model (model 4) [21] was created using the Statistica® Neural Networks 9.0 software to compare its robustness with that of the neural network models. The k-NN model was developed with the same descriptors and training cases as the neural model. The validation and test cases together formed the k-NN test subset. The input values were subjected to standardisation. The optimal number of neighbours was determined with a 10-fold cross-validation method, which evaluated k-NN prediction accuracy for k in the range of 1–40 (Fig. 2). The procedure randomly divided training cases into 10 subsets and evaluated its performance for a given subset based on the rest of cases. The accumulated averaged prediction accuracy provided a measure of model robustness for a given k value. Different types of distance metrics were assessed (such as Euclidian, Euclidean squared, City-block, and Chebyshev distance metric), but the City-block distance was the most efficient.

The quality of the neural and k-NN models was expressed as the percentage ratio of correctly predicted cases to the number of all cases in the test group.

3. Results

Model 1 (MLP 17:30-8-1:1) utilised 17 out of 21 initial input descriptors. The selected variables described the results of preoperative audiomteric tests, age, gender, the presence of ear discharge (infection), the extent of damage to the ossicular chain due to the disease, the type of pathological changes found in the middle ear, and the surgical procedure (approach, type of ossicular chain reconstruction, materials used for tympanic membrane and ossicular chain reconstruction, type of mastoid process surgery). The analysis of the network performance suggested overtraining, as the prediction quality for the training set was over 20% better than for the test set (see Table 2).

Model 2 (MLP 16:28-8-1:1) used the same input variables as model 1 with the exception of the surgical approach. Removal of one input parameter and retraining of the network increased the quality of prediction in the training and validation sets but did not influence the prediction accuracy in the test set.

Model 3 (MLP 16:38-9-2:1) used the input vector of model 2. Increasing the size of the hidden layer and modifying the output neuron activation function resulted in further improvement in the robustness of the network. Although the number of correct predictions in the training set was still higher than in the validation and test sets, the prediction quality in the test and validation set are in the same range (89% vs. 84%). These data indicate a higher generalisability of model 3 compared with the previous models.

Model 4 (the k-NN model) exhibited a similar prediction quality (47 correct predictions in test and validation sets) when compared to the first two neural models (48 and 50 correct predictions for models 2 and 3 in a combined test and validation set) and poorer prediction than the final model 3 (54 correct predictions). The k-NN model utilised the same input vectors as model 3 and used 1-nearest neighbour cluster and the City-block distance metric. Although a 10-fold cross-validation procedure showed that it is possible to achieve higher cross-validation quality for number of near neighbours in a range of 33–38 (Fig. 2), the resulting higher prediction quality of the combined test and validation subsets (in the range of 79%) arose from the uniform classification of all cases into an 'improvement' category. In these cases, the higher prediction quality was simply a statistical artefact.

4. Discussion

The quality of the neural networks developed in this study indicates that hearing improvement after tympanoplasty is, to some extent, predictable. The superiority of neural models (84% correct predictions in the test set, 89% correct prediction in the validation set) over the k-NN method (76% correct predictions in the evaluation set) suggests that the relationship between clinical variables and surgery results is complex and nonlinear. However, it should be underlined that the k-NN memory-based model 4 performed only a little worse than initially obtained neural models.

An important result of this study is the optimisation of the input vector and selection of the variables that are required for
prognosis. However, this study, and our previous experience in the field [22], shows that the problem of post-surgical hearing-outcome predictions is not trivial. Predictive models should always be constructed with caution, and in the context of ear surgery, there are several specific pitfalls (vide infra) that should be taken into account.

4.1. Data partitioning

To create a predictive model, the data need to be partitioned into at least two subsets. One subset is used for training, and the other is used for testing. In neural networks, a third subset of data is typically selected for model validation during learning to avoid overfitting.

4.1.1. The test set

The results obtained for the training and, to some extent, for the validation set are not informative of the model’s robustness. Only the quality of prediction in the test set is a measure of its generalisation ability and applicability to new cases [23].

If the available data set is small, the selection of a representative test set requires special attention. In experimental research small data sets are very frequent and data partitioning to training, validation and test sets significantly diminishes the number of training cases.

There is no analytical rule determining the exact number of training cases required for a neural model. The widely accepted rule of thumb suggests that the ratio of number of cases to the number of free parameters (weights) should be between 1 and 3 [24]. If the ratio value is below 1 (as in our case: 88 training cases/257 weights = 0.34) the model is prone to overfitting and special care should be taken to avoid it. However, even inclusion of the whole data set into the training subset would not solve the problem (ratio 150/257 = 0.58, which is still below 1). Therefore a rigorous external validation with a representative test set seems to be a reasonable trade-off.

There are many methods of data partitioning to select a representative test set. One of the most popular options in neural network software is random partitioning. Our previous experience [22] showed that, in a small sample of a highly non-homogeneous population, the random partitioning method frequently leads to the selection of a test set that is non-representative. Moreover, the test set may contain too many “simple” cases that are easily predicted by the model, even if the model is non-optimal (e.g., overfitted). As a result, the test set may falsely confirm the robustness of a model that, in reality, has very poor generalisation capability.

The method presented in this paper consists of data k-mean clustering prior to partitioning. Subsequently, every cluster is proportionally represented in the training, validation and test sets. This type of selection protocol, although laborious, is often superior to a random partitioning of the entire data set. This protocol provides a reliable method of model testing for smaller data sets without the necessity of cross-validation (such as n-fold cross-validation) or bootstrapping.

4.1.2. The training set

Random partitioning of a small number of cases may produce a training set where some subclasses are not represented, so the model is unable to learn how to produce correct outcomes for these subclasses. Therefore, it is vital to include all subclasses of cases in the training set. Data clustering and selection of training cases proportionally from every cluster ensures that the model will be able to accommodate its weights to the dataset that is representative for the whole studied problem.

Summing up, the clustering method can provide an optimal test and training set to create an effective and non-overfitted model for a dataset that could otherwise be considered too small for a given network.

4.2. Variable coding

In clinical data modelling, the use of arbitrary thresholding for continuous variables should generally be avoided [25]. In particular, the arbitrary categorisation of input variables can lead to unreliable results (as discussed by Aitchison et al. [26] in relation to different prognostic categories assigned to patients with a 3.9 mm and 4.0 mm thick melanoma). Variable values close to the cut-off point should be carefully analysed.

In artificial neural networks, it is not necessary to represent continuous input variables as categorical variables, but in our research, we were unable to avoid categorising the output variable. Numerous attempts to omit this problem were thus far unsuccessful. Because these models failed to predict hearing improvement after tympanoplasty for the studied group, they were not described in detail in this paper. These unsuccessful attempts included the following:

(i) the prediction of the exact value of the air conduction threshold by regression neural networks with a continuous output variable
(ii) the classification of the output variable into three categories (hearing improvement, no significant change or hearing deterioration) instead of two categories (hearing improvement versus lack of improvement).

Although we obtained satisfactory results using an arbitrary cut-off point for hearing results, it is crucial to modify this approach in further research concerning audiometric results. The precision of audiometry is limited, and its test–retest reliability depends on the quality of the audiometer, the experience of the audiometrist and the cooperation of the patient (because the patients need to signal when they hear a sound). The error for a single measurement can be as high as ± 10 dB. In comparison, the expected shifts in auditory thresholds after surgical procedures aiming exclusively at hearing improvements typically range between 10 and 45 dB in diseases other than CSOM, e.g., in otosclerosis, where hearing restoration is the only goal of treatment [27]. Therefore, it would be beneficial to interpret post-operative changes in hearing thresholds that do not exceed 10 dB as negligible [14,28]. To reduce the measurement error in our study, we used the mean value of three measurements conducted by one experienced technician using well-calibrated equipment.

4.3. Study group size

To avoid overfitting, it is important to maintain a reasonable balance between the network complexity and the number of training cases. A network composed of too many neurons has such an abundance of synaptic weights that it is not likely to develop sufficient generalisation capability. There are several empirical rules that help to estimate the optimal number of training cases for a given network [24]. Generally, more complex networks require larger data sets. If the available data set is limited, the size of the network should be diminished by reducing the number of input or hidden neurons. However, a network that is too simple may not be capable of solving a complex problem.

In the models presented in this study, several methods were used to avoid overfitting. The input vector dimensionality and number of hidden neurons was reduced until further reduction caused significant deterioration of the robustness of the network.
A comparison of the network’s predictive quality for the training, validation and test sets suggests that the model is not significantly overfitted. However, a network of this size will require a larger data set to ensure reliable generalisation quality in future research.

4.4. Study group heterogeneity

The clinical course of chronic suppurative otitis media is diverse, and many tympanoplasty techniques were developed to correct the various pathological changes in the middle ear. Our previous research [22] showed that creating separate models for more homogeneous subgroups of patients gives promising preliminary results. The subgroups were selected according to the extent of ossicular chain destruction and the type of surgical reconstruction. However, the size of the training sets in the subgroups was limited to the extent that the networks were prone to overfitting; these preliminary results should thus be considered with caution.

4.5. Generalisation quality

In this study, the test set was derived from the same population of patients as the training and validation set. A high testing quality suggests that the model can provide reliable predictions for new cases from the same population, i.e., for patients who undergo surgery in the Department of Otolaryngology in Krakow in the future. Because the surgical methods used for chronic suppurative otitis media in various hospitals are not uniform, only multicenter research will ensure a model’s universal generalisation quality.

4.6. Objective limitations in postoperative hearing improvement prediction

No prognostic model of postoperative hearing improvement in tympanoplasty can be expected to predict long-term results with absolute accuracy. Clinical observations show that the preoperative condition of the ear, the intraoperative findings and the surgical procedure are not the only factors that determine the final outcome. An unexpected trauma or infection may lead to postoperative destructions in the reconstructed tympanic cavity; in these cases, hearing thresholds can be much higher than expected. Information about the follow-up course cannot be implemented in a model that is designed to predict the outcome in advance (i.e., immediately after surgery). Postoperative complications were rare in the current study but are always possible. Prognostic models that report a prediction accuracy of 100% should be carefully examined to exclude the possibility of overfitting or improper testing.

4.7. Single time-point models versus multiple time-point models

The models described above predict hearing results at a single time point after surgery (12 months postoperatively). However, in some patients, hearing improvement after surgery is not stable over time. Further research in the field should focus on predicting patterns of postoperative hearing improvement and/or deterioration at multiple time-points. This task is similar to plotting survival curves in survival analyses [29].

5. Conclusions

As presented above, hearing improvement after middle ear surgery in chronic suppurative otitis media can be satisfactorily predicted by neural models. The clinical factors that influence the outcome are numerous; thus, modelling requires a detailed clinical description of the ear pathology and the surgical procedure, which are reflected by multidimensional input vectors. Because the pathological changes in the ear and surgical techniques in CSOM are very diverse, an effective prediction model requires an abundant training set. The problems encountered during model optimisation are not only those characteristic for neural networks in general but also those caused by limitations in audiometry in providing reliable hearing thresholds values. Because of the potential and unpredictable complications in post-surgical follow-ups, no model can achieve absolute accuracy in postoperative hearing improvement predictions. Nevertheless, neural models seem to be better than other approaches considered by the authors and discussed in the literature; these neural networks can, therefore, be recommended for practical use.

6. Summary

The paper presents a method of development of predictive neural models for hearing improvements after middle ear surgery for chronic otitis media.

The purpose of this paper is not only to present a solution to a particular medical issue but also to describe how to process clinical data for neural network learning and interpret output signals to answer medical questions. The studied data set comprised 150 patients that were partitioned into three subsets: learning, validation and test based on the cluster method. Each patient was characterised by the following set of input variables: age, gender, preoperative audiometric results, ear pathology and details of the surgical procedure. The predicted (output) variable was the postoperative hearing threshold. The predictions were performed using artificial neural networks and k-nearest neighbour (k-NN) models. To obtain satisfactory robustness of the models, a carefully optimised multidimensional input vector was necessary. The optimisation of input vector was performed by heuristic algorithm (Intelligent Problem Solver) implemented in the Statistica Neural Network module followed by manual optimisation of neural networks based on post-processing sensitivity analysis.

The best neural model developed in this study (MLP 16:38-9-2:1) achieved 98%, 89%, and 84% correct predictions for the training, validation, and test data set, respectively, while the k-NN model produced only 75.8% correct predictions for cases belonging to validation and test sets. The superiority of neural over the k-NN method suggests that the relationship between clinical variables and surgery results is complex and nonlinear.

Several additional issues were also addressed, such as methods of data partitioning, problems involved in variable coding (especially a problem of the arbitrary categorisation of variables), effect of a patient group size and heterogeneity on the model prediction and generalisation quality, as well as a need for transduction from a single time-point models toward multiple time-points models in the modelling of such phenomena as postoperative health improvement.

Our studies proved that hearing improvement after middle ear surgery in chronic otitis media can be satisfactorily predicted by neural models. Other tested methods proved to be less effective in postoperative hearing predictions.

Conflict of interest statement

The authors declare no conflicts of interest.
References


