Methodology of Constructing Statistical Models for Nonlinear Non-stationary Processes in Medical Diagnostic Systems

Peter Bidyuk [a], Irina Kalinina [b] and Aleksandr Gozhyj [b]

[a] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*
[b] *Petro Mohyla Black Sea National University, Nikolaev, Ukraine*

Abstract
The article presents a methodology for analysis and modeling nonlinear and non-stationary processes associated with medical diagnostics and solving medical problems. The methodology is based on collecting and preliminary processing of statistical data, identifying and accounting for possible uncertainties, building and estimating the structure of mathematical model, and evaluating its parameters, estimating model based forecasts and calculating statistical criteria for the model adequacy as well as quality of the forecasts. The analysis of selected models for linear and nonlinear processes is presented. A scheme for combining forecasts when estimating a diagnosis is proposed. An example of using a diagnostic system for predicting a patient's condition using combined (linear + nonlinear) model is given and the methods used are analyzed.

Keywords 1
Medical decision support systems, Nonlinear and non-stationary processes, Statistical model, Preliminary processing, Uncertainties, Combining forecasts.

1. Introduction

Appropriately designed medical decision support system (DSS) can provide a substantial help regarding estimation of medical diagnosis, analyzing the processes taking place during the period of patient treatment, forecasting patient state, modeling complex processes and situations so that to derive correct conclusions, perform necessary simulations, generate appropriate advices etc. [1]. DSS can easily analyze available and newly coming statistical data and expert estimates regarding various processes taking place in medical environment. DSS can perform in a very short time sophisticated and cumbersome computations, and provide final results in appropriate, convenient for medical staff form. The well-known example of successfully working medical DSS is Quick Medical Reference (QMR) diagnostic system [2, 3] based upon Bayesian, neural, statistical, and other methods for data analysis and generating probabilistic inference regarding diagnosis, situational analysis, drawing conclusions etc.

Some of the process studied in medical applications, are nonstationary or piecewise stationary and contain nonlinearities. It means that their statistical parameters may change in time what requires special attention regarding modeling, analyzing and forecasting the processes being studied. The processes are also characterized by availability of stochastic or deterministic trends (conditional expectation varies in time) dependently on specific situations, random disturbances and factors influencing them. Usually nonstationary processes exhibit nonlinearities of various kind (nonlinearity regarding variables or parameters). The deterministic trend regarding patient state can be formally described by the linear, quadratic or cubic function, exponent, spline or harmonic function.

Estimation of variance is also an important stage in data research in medical diagnostic systems [4 - 6]. This is a key statistical parameter for generating correct diagnosis based on available statistical

data and forecasts of relevant conditional variance. One of important points in modeling dynamics of various processes is identification and taking into consideration possible uncertainties related to the data available and expert estimates. The uncertainties are considered here as the factors of negative influence to the modeling process as a whole that result in various computational errors decreasing quality of intermediate and final results. As an example of possible uncertainties could be mentioned measurement errors produced by the clinical laboratory devices, and previously unknown consequences resulting from prescribed medications. Another problem is created by the so called structural uncertainties related to estimation of model structure.

This study is directed towards further refinement of mathematical model constructing methodology for nonlinear non-stationary processes in medical applications to be further used in specialized diagnostic decision support systems. It is touching upon improvement of data quality before model constructing, as well as model structure and parameter estimation techniques using alternative statistical data analysis procedures.

Problem statement. The paper is focused on solving the following problems: development of systemic methodology for mathematical models constructing for linear and nonlinear process in medical applications; development of decision support system structure and its functions for application in medical diagnostics; providing a review of some mathematical models for possible prospective applications in medical decision support systems; presenting an example for possible DSS application.

2. Materials and Methods

This section discusses the methodology for analysis and modeling of nonlinear and non-stationary processes associated with medical diagnostics and solving medical problems. The analysis of selected models for linear and nonlinear processes is presented. A possible scheme for combining forecasts when making a diagnosis is proposed. An example of using the diagnostic system for prediction of a patient state and the efficiency of the data analysis methods are proposed.

2.1. DSS structure and functions

The DSS proposed (Fig. 1) includes the following functional blocks: user interface, central computing subsystem generating required results of data analysis according to specific problem statement, knowledge and database (KDB), as well as intermediate and final results representation subsystem. The KDB contains all necessary computational procedures, sets of a model and forecasts quality criteria, statistical data, relevant expert estimates and best model selecting rules. All computations are performed within the central computing subsystem that generates intermediate and final results of data analysis according to the user requests. The results representation subsystem provides a user with necessary information regarding the computational procedures, and data presentation in convenient formats.

The basic functions of the system are as follows: data collection from local and external sources; pre-processing, i.e. preparing data for model constructing and subsequent processes forecasting; model structure and parameter estimation; computing forecasts of patient state and combination of the separate forecasts; generation of recommendations and diagnostic messages; retrospective analysis of available former results (for example, retrieval from memory and analysis of available diagnostic messages) etc. The functional possibilities of the system are easily modified and expanded thanks to the modular system construction. At the core of the diagnostic subsystem are intellectual data analysis procedures such as Bayesian networks, neural networks and fuzzy-neural models, decision trees etc. The selected statistical procedures are also widely used for preliminary data processing, identifying and fighting statistical uncertainties, correlation and variance analysis of data, estimation of resulting models adequacy, and quality of the model based forecasts; generation of alternative decisions based upon probabilistic and statistical procedures.
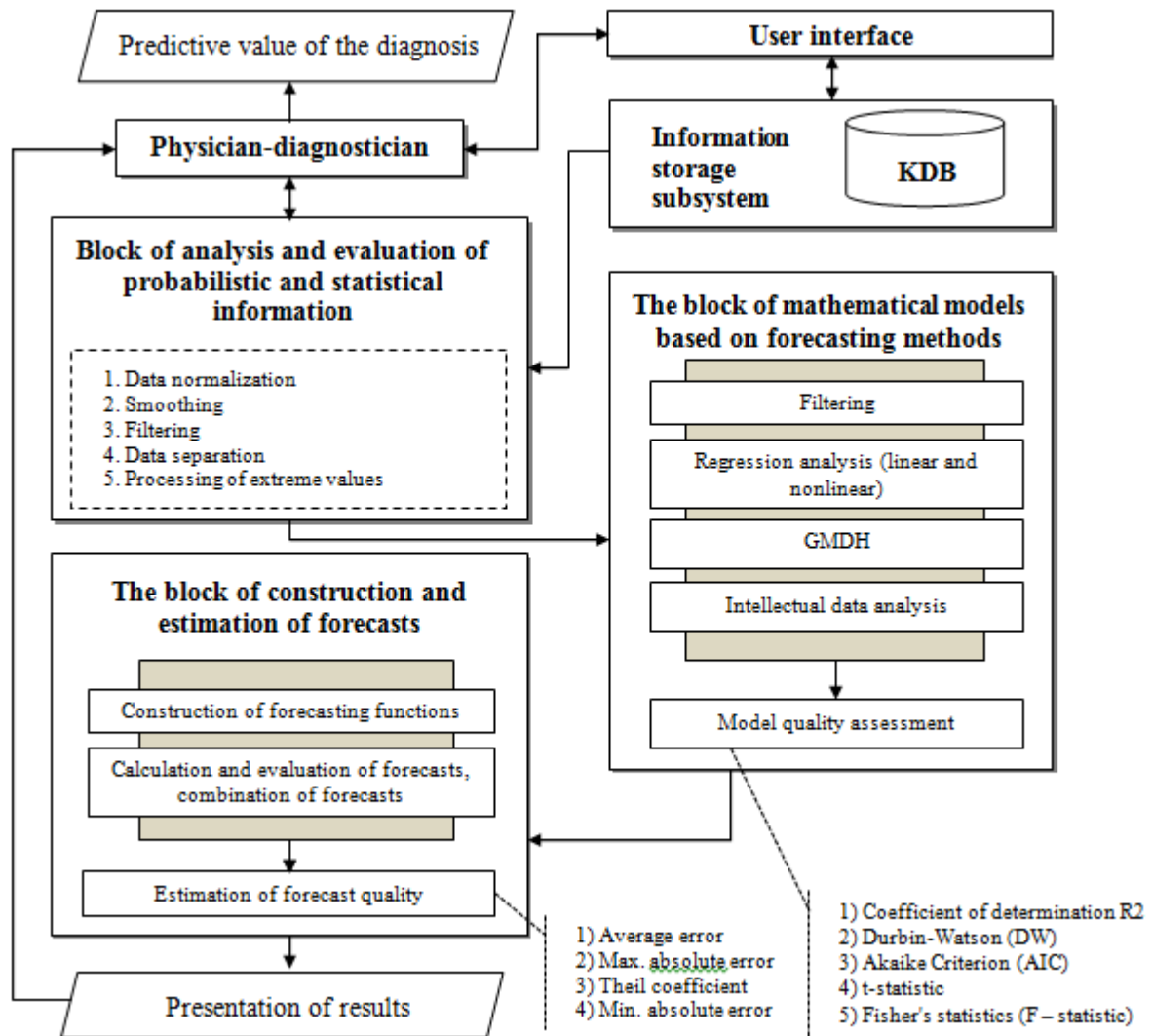
Figure 1: DSS structure

2.2. The methodology of modeling proposed

The methodology for modeling and analyzing medical data in solving diagnostic problems is based on the following steps:

- Collection and preliminary statistical processing of data for building a model (adding missing data, normalizing data, filtering, accounting and processing possible outliers, etc.).

- Identification, estimation and accounting for uncertainties in the data (estimation of data values that cannot be measured, estimation of statistical parameters for observations (mean, median, variance, covariance, etc.); determination of the required data structure; analysis of statistical characteristics for the random samples (type of distribution and its parameters) that may affect for the diagnosis.

- Evaluation of mathematical models structure using statistical and probabilistic methods of data analysis for the selection and further use of the best of them in solving the forecasting and decision making problems. The following parameters are used as the main characteristics of the model structure being constructed: the model dimension (total number of equations that form the model); model order (differential equation order, or autoregression and the moving average orders); the presence of nonlinearity of the process and estimation of its type (nonlinearity in variables and / or in terms of model parameters); estimation of the input time delay (lag) of the process, etc.

To solve the problem of identification and taking into consideration possible nonlinearities it is recommended to construct separately the models for linear and nonlinear parts of a process being studied using various possibilities (special nonlinear components) for formal description of the nonlinear part. The acceptable quality results were achieved with application of combined linear and nonlinear regression; linear regression and neural or Bayesian networks; linear regression and special nonlinear functions like polynomials and nonparametric kernels etc.

The following methodology is proposed for determining and describing possible nonlinearities of the data being analyzed:

- Correct estimation of model parameters based on alternative methods (NLS, ML, MCMC and other methods) allows for calculating unbiased parameter estimates, determining the type of distribution for variables and its parameters, estimating alternative structures of the model. The use of such alternative estimation methods in decision-making procedures allows for further comparison of estimates and the selection of the best model.
- Computing statistical parameters that are quality criteria and characterize adequacy of mathematical models. On their basis, the most adequate model is selected. Usually there exist several candidate models that can be constructed using different methods. The final choice of the model is performed after its application to solve the specific problem stated.
- Construction of forecasts (usually short-term or medium-term) and evaluation of their quality to determine the best predictive model. For this purpose the following statistical quality criteria are used: Theil coefficient, MAPE (mean absolute percentage error), MAE (mean absolute error), etc.
- Testing the models constructed using the processes data with similar statistical characteristics (model calibration).

Usually in the model constructing procedures it is necessary to take into account the following types of uncertainties: data uncertainties, model structure uncertainties, and parametric uncertainties. The data uncertainties include: missing measurements, the presence of short uninformative data samples, possible extreme values (outliers), distortions due to availability of observation and external noise processes, etc. Basically, these types of uncertainties are easily handled using filtering procedures. Uncertainties regarding the structure of the model are due to poor data structure, which does not contain enough necessary information for estimating the model structure and its parameters. Parametric uncertainties depend on the quality of the statistical data available. Usually they are observed in the form of biased parameter estimates. Elimination of the bias is performed by applying several methods of parameter estimation, such as: OLS, maximum likelihood (ML) algorithms, and Monte Carlo algorithms with Markov chain [6, 7].

To monitor the software implementation of the methodology it is necessary to consider at least three sets of statistical quality criteria during the modeling process: data quality parameters, model adequacy, and forecast quality statistics. It is necessary to provide for the software implementation of the methodology for the analysis of alternative solutions developed on the basis of the calculated forecasts.

The practical application of the proposed modeling methodology lies in solving the problem of predicting the patient's condition, monitoring this condition, as well as making decisions regarding diagnostics, and modeling complex situations with the use of specific computational procedures.

2.3. Some widely used models of linear and nonlinear processes

Today, there are many statistical methods for analyzing and studying the processes associated with clinical treatment and diagnosis identification based upon regression analysis results. One of the approaches is based on the state space (SS) representation of the models constructed. Based on this approach, the patient's condition is assessed and predicted [8 - 14]. Another approach to developing linear and nonlinear models for diagnosis is based on data mining (IDA) and machine learning (ML) methods like following: neural networks, fuzzy sets, neural fuzzy models, Bayesian networks (static and dynamic), complex multivariate probability distributions, models describing interactions of various factors, non-parametric and semi-parametric models, decision trees, etc.

The structure of a time series mathematical model can be described by the following expression:

$$S = \{r, p, m, n, d, w, l\},$$

where r is model dimension (number of model equations); p is model order (maximum order of differential or difference equation used for specific process description); m is a number of independent variables (regressors); n is nonlinearity and its type (with respect to variables or parameters); d is input delay time (lag or medication transport delay); w is external (or possibly internal) stochastic disturbance and type of its probability distribution; l represents possible constraints on variables and/or parameters [15].

In medical practice it is possible to use the models based on nonlinear regression. For example, similar models can be used to describe two interrelated processes, $y_1(k)$, and $y_2(k)$ when predicting the patient's condition and identifying diagnosis:

$$y_1(k) = a_0 + a_1 y_1(k-1) + b_{12} \exp(y_2(k)) + a_2 x_1(k) x_2(k) + \varepsilon_1(k),$$
$$y_2(k) = c_0 + c_1 y_2(k-1) + b_{21} \exp(y_1(k)) + c_2 x_1(k) x_2(k) + \varepsilon_2(k),$$

where $y_1(k)$ is a principal state variable for the first process under study; $y_2(k)$ is a principal state variable for the second process; $x_1(k)$ is a level of the first medication being used; $x_2(k)$ is a level of the second medication. Sometimes, it is also possible to apply the following generalized linear model:

$$y(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j v(k-j) + \sum_{i=1}^{m} \sum_{j=1}^{s} c_{i,j} y(k-i) v(k-j) + \varepsilon(k),$$

where p, q, m and s are positive numbers that represent the model order [15].

The complete model of nonlinear processes can be based on linear combination of linear and nonlinear components like follows:

$$y(k) = \beta^T \mathbf{z}(k) + \sum_{i=1}^{p} \alpha_i \varphi_i \left(\theta_i^T \mathbf{z}(k) \right) + \varepsilon(k),$$

where $\mathbf{z}(k)$ is a vector of time delayed values of basic dependent variable $y(k)$, as well as former and current values of independent explaining variables $\mathbf{x}(k)$ with appropriate values of time delay. Here, $\varphi_i(x)$,, is a set of linear and possibly nonlinear functions that may include the following components: power function $\varphi_i(x) \equiv x^i$; the harmonic trigonometric functions like $\varphi_i(x) = \sin x$ or $\varphi_i(x) = \cos x$, etc. If necessary, this equation can be expanded with quadratic form of the type; $\mathbf{z}^T(k) \mathbf{A} \mathbf{z}(k)$; $\varphi_i(x) = \varphi(x)$, $\forall i$, where $\varphi(x)$ is a suitable link function, for example appropriate probability density function or logistic function of the type:

$$\varphi(x(k,z)) = \frac{1}{1 + \exp(-x(k,z))},$$
$$x(k) = \alpha_0 + \alpha_1 z_1(k) + \cdots + \alpha_m z_m(k) + \varepsilon(k),$$

where $z_i(k), i = 1, 2, \ldots, m$ are explaining variables for the intermediate principal variable $x(k)$, and $\varphi(x(k,z))$, respectively.

Another general class of nonlinear models (suitable for modeling and forecasting patient state) can be presented as follows:

$$\mathbf{y}(k) = \sum_{j=1}^{p} \varphi_j(\mathbf{x}(k-1)) \mathbf{y}(k-j) + \mu(\mathbf{x}(k-1)) + \varepsilon(k),$$

where $\mathbf{y}(k)$ is $[n \times 1]$ vector of dependent variables; $\mathbf{x}(k) = [\mathbf{y}(k), \mathbf{y}(k-1), \ldots, y(k-n+1)]$ is a vector of state variables; here dynamics of the state variables can be described by the following state space model:

$$\mathbf{x}(\mathbf{k}) = h(\mathbf{x}(k-1)) + \mathbf{F}(\mathbf{x}(k-1)) \mathbf{x}(k-1) + v(k).$$

When constructing patient state forecasting model usually there are developed several candidates with subsequent selection of the best one using a set of model adequacy criteria, say determination coefficient, Student t-statistics, Durbin-Watson statistics, Akaike information criteria and others suitable for specific case. The separate criteria can be used for constructing a single combined criteria that enables for automatic selection of the best model.

Bayesian networks. Mathematical models in the form of Bayesian networks (BN) are very useful for creating clinical diagnostic systems as well as diagnostic systems for engineering and economic applications. Generally BN is defined as a directed acyclic graph the vertices (nodes) of which are

variables selected to characterize behavior of a system (or multivariate processes) under study, and the arcs indicate to existing causal relations between the variables. To each daughter node of BN is assigned conditional probability table that is used for computing probabilistic inference. The model has the following advantages: dimension of the model can be very high, and the variables hired can be discrete or continuous; the separate facts generated by experts can also be taken into consideration. Today there exist numerous procedures for estimating BN structure and computing probabilistic inference on their basis. Besides, the models are suitable for fighting probabilistic uncertainties of the type: "is the event going to happen or not, and what is the probability of occurring the event?" The known successful applications of BN today are numerous and their number continues to grow.

Generally the BN constructing procedure includes the following steps:
- research problem statement;
- a thorough analysis of a system (processes) under consideration aiming to revealing specific features of its functioning, as well as selection of parent and daughter variables;
- identification of existing system models and determining the possibilities for their usage in the frames of the DSS constructed;
- estimating existing causal relations between the variable selected using appropriate set of statistics;
- possible reduction of the model dimensionality;
- scaling and (possibly) discretization of the model variables selected;
- determining semantic (logical) constraints for the model;
- structure estimation for candidate models using appropriate optimization procedures and model quality criteria;
- model adequacy analysis and selection of the best one(s);
- application of the model(s) constructing to solving the problem stated (step 1); comparison of the results obtained with other possible models;
- the final model selection.

Structural uncertainties. According to the methodology proposed the following procedures are used to cope with possible structural uncertainties of a model: refinement of model order by applying recursive adaptive approach to modeling and automatic search for the "best" structure using combined statistical criteria; adaptive estimation of input delay time, and the type of statistical data probability distribution with its parameters; describing detected nonlinearities with alternative analytical forms, and with subsequent quality estimation of the forecasts generated [12, 13]. Another wide class of nonlinear heteroscedastic processes exists today, and is described by the models of conditional variance dynamics. Usually studying of such processes includes constructing of the two model types: model for the process amplitude, and the model for time changing variance. As far as formal description of variance is based on quadratic variables and functions, heteroscedastic processes are nonlinear by definition [11].

An important stage in the constructing of statistical diagnostic models, including predictive ones, is the assessment of their quality. Usually, two or more state prediction methods are used to compute forecast scores in order to be able to combine predictions to further improve the state prediction score. The forecasts combining scheme with equal or different weighting coefficients used in the approach proposed is explained by Fig. 2.

The model based forecasts can be computed, for example, with six selected techniques as shown in Fig. 2. Regression model (autoregression (AR) or AR with moving average (ARMA)) is used for generating forecast as well as its transformed version into state space (SS) form is necessary for further application of optimal Kalman filter (KF). Adaptive version of KF is interesting from the point of view that it provides a possibility for forecasting and on-line (or off-line) estimation of state disturbance and measurement noise covariance. An alternative approach to Kalman filtering is application of Bayesian probabilistic (particle) filter providing for the forecasts in the form of probability distribution that can be selected of necessary type. The distribution can be further used for estimating its parameters showing future patient state as well as possible span of the state values in selected space.
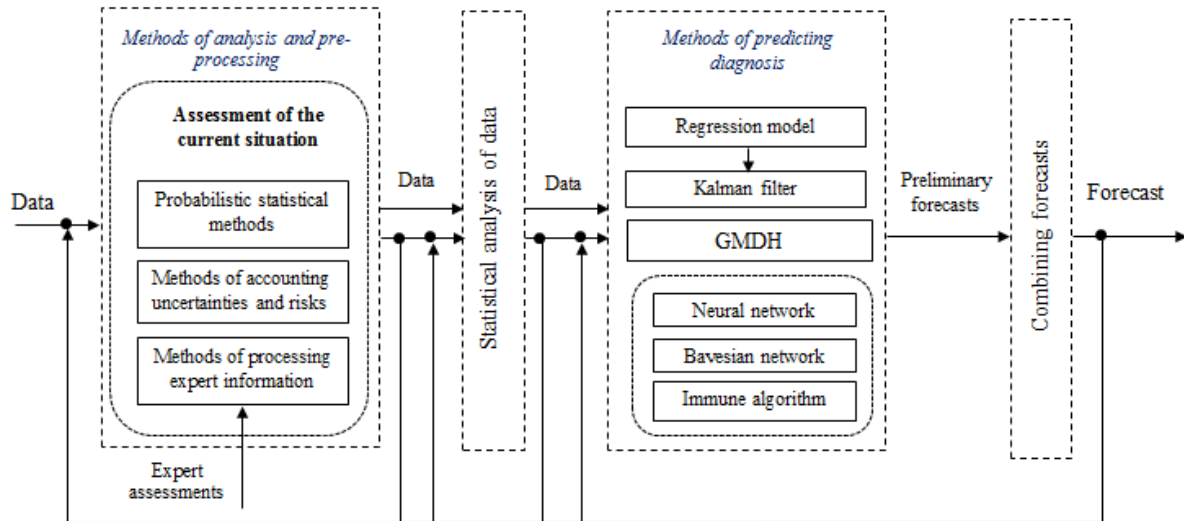
Figure 2: The principle of combining alternative forecasts

Some other advantages of using the probabilistic approach to filtering are as follows: it helps to take into consideration some uncertainties relevant to conditional probabilities of events hidden in the data being processed; and it helps to generate possible future paths for the states under consideration using appropriate mathematical models. Also the Bayesian filtering approach represents useful instrument for data processing and events simulation in parallel with optimal Kalman filter on the purpose of comparison of the results achieved and computing alternative state estimates. Generally the diagnostic system under development should contain a set of digital, optimal and probabilistic filters providing the possibilities for filtering and smoothing statistical\experimental data, imputation of missing measurements, preform short term prediction of states and estimate possible available non-measurable variables or parameters.

The well-known group method of data handling and modeling procedure (GMDH) provides the possibility for constructing models in the general form of Kolmogorov-Gabor polynomials, and the last three methods mentioned in the figure are related to the popular today intellectual data analysis techniques. The GMDH approach to modeling is very convenient from the point of view that it estimates model structure by its internal procedures.

Thus, here we propose the combination of classic regression (statistical) approach with the intellectual data analysis methodology. The best result of combining the forecasts with respect to enhancement the quality of final forecast is achieved when variances of forecasting errors for selected forecasting techniques do not differ substantially (say, not by an order). Some other possibilities for hiring possible linear and nonlinear models to describe and forecast patient state are given in the Table 1 below.

The structure of models No. 1-8, presented in Table 1, is partially determined and can be changed (refined) in the process of adaptation using specific statistical data. Model 1 can be used to describe various trends along with deviations from the conditional mean. Models 2 and 4 describe bilinear and exponential nonlinearity. Model 3 describes nonlinearity with saturation. Models 5 and 6 are used to describe the changes in conditional variance in the study and modeling of heteroscedastic processes. Model No. 6 shows the best results for short-term forecasting of conditional variance. Models 7, 8, and 9 can be used to describe arbitrary nonlinearities with high-order model members. The use of fuzzy sets in modeling involves the construction of a set of rules that describe processes or systems and carry out logical inference under conditions of information uncertainty. The models based on neural networks and fuzzy neural networks are used to model complex nonlinear functions under conditions where some variables are unobservable. Bayesian networks (static and dynamic) are statistical and probabilistic models, with the help of which it is possible to model complex multidimensional processes with obtaining the final result of their application in the form of probabilistic inference (conditional probabilities) characterizing the patient's condition [16].

Table 1

Some linear and nonlinear models for describing process dynamics

No.	Model description	Formal model structure
1	AR + polynomial of time	$(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + b_1 k + \cdots + b_m k^m + \varepsilon(k),$ $k = 0,1,2,\ldots$ is discrete time; $t = kT_s$; T_s is sampling time.
2	Generalized bilinear model	$y(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j v(k-i) =$ $= \sum_{i=1}^{m} \sum_{j=1}^{s} c_{ij} y(k-i) v(k-j) + \varepsilon(k)$
3	Logistic regression	$\varphi\big(x(k,z)\big) = \dfrac{1}{1 + \exp(-x(k,z))},$ $x(k) = \alpha_0 + \alpha_1 z_1(k) + \cdots + \alpha_m z_m(k) + \varepsilon(k)$
4	Nonlinear extended econometric autoregression	$y_1(k) = a_0 + a_1 y_1(k-1) + b_{12} \exp\big(y_2(k)\big) + a_2 x_1 x_2 + \varepsilon_1(k),$ $y_2(k) = c_0 + c_1 y_2(k-1) + b_{21} \exp\big(y_1(k)\big) + c_2 x_1 x_2 + \varepsilon_2(k)$
5	Generalized autoregression with conditional heteroscedasticity (GARCH)	$h(k) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^{p} \beta_i h(k-i)$
6	Exponential generalized autoregression with conditional heteroscedasticity (EGARCH)	$\log[h(k)] = \alpha_0 + \sum_{i=1}^{p} \alpha_i \dfrac{|\varepsilon(k-i)|}{\sqrt{h(k-i)}} + \sum_{i=1}^{p} \beta_i \dfrac{\varepsilon(k-i)}{\sqrt{h(k-i)}} +$ $+ \sum_{i=1}^{q} \gamma_i \log[h(k-i)] + v(k)$
7	Nonparametric model with functional coefficients	$y(k) = \sum_{i=1}^{p} \big\{ \alpha_i + \big(\beta_i + \gamma_i y(k-d)\big) \cdot \exp\big(-\theta_i y^m(k-d)\big) \big\} +$ $+ \varepsilon(k)$
8	Radial basis function	$f_\theta\big(x(k)\big) = \sum_{i=1}^{M} \lambda_i \exp\left(-\dfrac{(x(k) - \mu_i)^2}{2\sigma_i^2} \right) + \varepsilon(k),$ $\theta = [\mu_i, \sigma_i, \lambda_i]^T; M = 2,3,\ldots$
9	State-space representation	$\mathbf{x}(k) = \mathbf{F}[\mathbf{a}(k), \mathbf{x}(k-1)] + \mathbf{B}[\mathbf{b}(k), \mathbf{u}(k-d)] + \mathbf{w}(k$
10	Neural networks	Selected (constructed) network structures
11	Fuzzy sets and neuro-fuzzy models	Combination of fuzzy variables and neural network model
12	Dynamic Bayesian networks	Probabilistic Bayesian network structure constructed with data and/or expert estimates
13	Multivariate distributions	Say, copula application for describing multivariate distribution
14	Immune systems	Immune algorithms and combined models

2.4. Example of the medical DSS application

The example is touching upon patient state forecasting using combined (linear + nonlinear part) model. The combined model proposed includes optimal and digital filters, linear regression models and nonlinear logit model (Fig. 3).
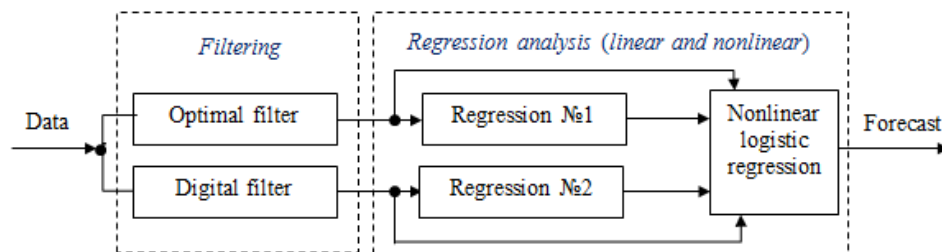

Figure 3: Combined model: filtering + linear regression + nonlinear regression

The purpose of using the two alternative filters is to perform data smoothing (suppressing undesirable high frequency components often contained in measurements) and this way prepare it for modeling and state forecasting. The two filters provide two alternatives for subsequent constructing linear regression models. Besides, application of the optimal Kalman filter additionally provides a possibility for solving the following problems: estimation of non-measurable state vector components; variance/covariance estimation for observations hired for model constructing; and short-term state forecasting when necessary. In this specific example the following hourly measurements were used: arterial blood pressure, heart rate, skin resistance, and body temperature. The model purpose was to provide a forecast for evolution of a patient state to "better" (indicated by "1"), and to "worse" (indicated by "0"). Table 2 contains quality of forecasting direction of the patient state evolution.

Table 2

Results of forecasting direction for evolution of patient state

Model type	Probability of correct direction forecast
Logistic regression	72.75%
Classification tree	69.61%
Logistic regression + extra forecast by linear model	77.69%
Classification tree + extra forecast by linear model	75.54%

Thus, in both cases (logistic regression and classification tree) the best state forecasting results were achieved with the use of extra state forecast by the linear regression model. The statistical quality characteristics of the forecasts achieved show high quality of the forecasts and possibility of their use in estimation of patient state. Certainly there are possibilities for further improvement of the preliminary results obtained.

3. Conclusions

The article proposed the technique for modeling and forecasting nonlinear non-stationary processes based on the analysis and processing of statistical data in medical applications. The technique is based on general systemic (system analysis) principles and represents a hierarchical structure in the analysis of medical data, taking into account possible statistical uncertainties. With the help of the systemic approach, the development of adaptive schemes for assessing the structure and parameters of the model, the use of statistical and probabilistic criteria for the construction and selection of the best model for solving medical problems is implemented. The procedures for optimal and digital filtering of data, methods for filling in the missing measurements, methods for estimating model parameters and Bayesian programming were proposed as methods for overcoming possible uncertainties.

On the basis of the methodology developed, it is possible to construct combined models, including statistical regression and probabilistic models in the form of Bayesian networks, decision trees, logistic regression, etc. This approach has shown its high effectiveness in constructing short-term forecasts. The presented application example demonstrated the adequacy of the model constructed and high quality of short-term patient states predictions.

Further development of the proposed methodology will be aiming at the development of more advanced model structures for nonlinear non-stationary processes encountered in medical applications to improve diagnostic and therapeutic processes. The technique is implemented in a diagnostic medical decision support system and can be used for short-term prediction of a patient's condition and diagnosing its state.

4. References

[1] W. Wojcik, A. Smolarz, Information Technology in Medical Diagnostics, CRC Press, London, 2017.

[2] J.B. Lemaire, J.P. Schafer, L.A. Martin, P. Faris, M.D. Ainslie, R.D. Hull, Effectiveness of the QMR as a diagnostic Tool, Canadian Medical Association Journal, 1999, Vol. 161, No.6, pp. 725 – 728.

[3] A. Linton, Quick Medical Reference, Bulletin of the Medical Library Association, 1990, Vol. 81, No.3, pp. 347 – 349.

[4] F.X. Diebold, Forecasting in economics, business, finance and beyond, University of Pennsylvania, 2015.

[5] B.E. Hansen, Econometrics, University of Wisconsin, 2017.

[6] R.S. Tsay, Analysis of financial time series, New York: John Wiley & Sons, Inc., 2010.

[7] S.O. Dovgij, O.M. Trofymchuk, P.I. Bidyuk, DSS based on statistical and probabilistic procedures, Kyiv: Logos, 2014.

[8] P. Mangiameli, D. West, R. Rampal, Model selection for medical diagnosis decision support systems, Decision Support Systems, 2004, Vol. 36, issue 3, pp. 247-259. DOI: 10.1016/S0167-9236(02)00143-4.

[9] B. Malmir, M. Amine, S. Chang, A medical decision support system for disease diagnosis under uncertainty, Expert Systems with Application, 2017, Vol. 88, pp. 95-108. DOI: 10.10161j.eswa.2017.06.031.

[10] P. J. Twomey,1 M. H. Kroll, How to use linear regression and correlation in quantitative method comparison studies, International Journal of Clinical Practice, Journal compilation 2008, Blackwell Publishing Ltd. Int. J Clin Pract, Vol. 62, No. 4, pp. 529–538. DOI: 10.1111/j.1742-1241.2008.01709.x.

[11] J.G. De Gooijer, Elements of Nonlinear Time Series Analysis and Forecasting, Cham (Switzerland): Springer, 2017.

[12] R. W. Alexandrowicz, R. Jahn, F. Friedrich, A. Unger, The importance of statistical modeling in clinical research, Neuropsychiatr (2016), Springer, Vol. 30, pp. 92–102. DOI: 10.1007/s40211-016-0180-3.

[13] P. Mishra, C. M. Pandey, U. Singh, A. Keshri, M. Sabaretnam, Selection of Appropriate Statistical Methods for Data Analysis, Annals of Cardiac Anaesthesia, 2020, Vol. 22, Issue 3, pp. 297-301. DOI: 10.4103/aca.ACA_248_18.

[14] N. Kim, A. H. Fischer, B. Dyring-Andersen, B. Rosner, G. A. Okoye, Research Techniques Made Simple: Choosing, Appropriate Statistical Methods for Clinical Research, Journal of Investigative Dermatology, 2017, Vol. 137, pp. 173-178. DOI:10.1016/j.jid.2017.08.007.

[15] O. L. Tymoshchuk, V. H. Huskova, P. I. Bidyuk, A combined approach to modeling nonstationary heteroscedastic processes, Radio Electronics, Computer Science, Control, 2019, No.2, pp. 80 – 89. Doi: 10.15588/1607-3274-2019-2-9.

[16] P. Bidyuk, A. Gozhyj, I. Kalinina, V. Gozhyj, Analysis of uncertainty types for model building and forecasting dynamic processes, Advances in Intelligent Systems and Computing II 689. Springer-Verlag, pp. 66-78. Doi: 10.1007/978-3-319-70581-1.