

---

**ПРИМЕНЕНИЕ ПРОЦЕССА ГАУССА И КЛАССИФИКАТОРА НЕЙРОННОЙ СЕТИ В ЗАДАЧЕ ПРЕДСКАЗЫВАНИЯ СЕПСИСА У ПАЦИЕНТОВ**

---

*Кузьмин Е.А.1,  
Качальский В.Г.2,\**

*<sup>1</sup> Новосибирский государственный технический университет,  
Новосибирск, Россия;*

*<sup>2</sup> Новосибирский государственный технический университет,  
Новосибирск, Россия;*

*\* Корреспондирующий автор*

**АННОТАЦИЯ**

Ранний прогноз сепсиса уменьшает показатель смертности у пациентов. В работе представлена система предсказания сепсиса, оперирующая неструктурированными, изменяющимися во времени данными о пациентах. С помощью многомерного процесса Гаусса моделируются непрерывные показатели пациента, сохраняя при этом степень неопределенности в каждый момент времени. Это позволяет учитывать нерегулярность заполненных данных, отображающих реальную клиническую картину. Далее данные поступают в классификатор, реализованный на основе рекуррентных нейронных сетей (РНН), которые хорошо справляются с большим количеством входящих показателей и способны обрабатывать серии событий с учетом времени. Исследование показало, что точность предсказания системы за двенадцать часов, четыре часа и один час равняется 81, 89 и 91 процентам соответственно.

**ABSTRACT**

An early and accurate sepsis prediction can significantly reduce mortality. The work presents a sepsis prediction system, which operates with unstructured, time-varying patient data. The multivariate Gaussian process simulates continuous patient indicators while maintaining the degree of uncertainty at each time. This allows to save irregularity of the filled-in data that displays the real clinical picture. Then the data is sent to a classifier implemented on the basis of recurrent neural networks (RNN), which cope well with a large number of incoming indicators and can process a series of time events. The study found that the system prediction accuracy in twelve hours, four hours, and one hour is 81, 89, and 91 percent, respectively.

**Ключевые слова:** процессы Гаусса, сепсис, нейронные сети, прогнозирование.

**Keywords:** Gaussian process, sepsis, neural network, prediction.

Сепсис – это смертельно опасное заболевание, которое возникает, когда бактериальная инфекция, поразившая какую-либо часть тела, попадает в кровоток. Ответ организма на эту инфекцию приводит к повреждению его собственных тканей и органов, что ведет к их отказу, а далее к смерти человека. Признаки и симптомы сепсиса у пациентов могут быть различными в разные моменты времени поскольку болезнь постоянно меняет свой характер. К симптомам относятся повышение или понижение температуры тела, озноб, изменение психического состояния, затрудненное/учащенное дыхание, учащенное сердцебиение, ослабление пульса, низкое кровяное давление и так далее.

Для того чтобы повысить скорость определения состояния пациента и улучшить лечение, были созданы критерии, которые отражают степени тяжести хронических заболеваний, отказа функций органов и остального состояния пациента в количественных оценках.

Критерий SIRS оперирует пограничными значениями четырех показателей: температура, частота сердцебиения, частота дыхательных движений и общий клинический анализ крови [1].

Но множество экспертов рассматривают SIRS оценку как слишком чувствительную, так как около 90 процентов пациентов, поступивших в отделение интенсивной терапии (ОИТ), попадают под SIRS критерий [2]. Поэтому для более точного прогнозирования был предложен новый критерий

SOFA [3].

Сепсис продолжает быть одним из лидирующих причин смертности среди пациентов ОИТ. Исследования показывают, что каждый час задержки антибактериальной терапии увеличивает вероятность летального исхода на 7 процентов, а через 36 часов начатая антибактериальная терапия практически не влияет на выживаемость пациента [4]. Ранний и точный прогноз возникновения болезни может значительно снизить уровень смертности, а также снизить риск повреждения органов. Поэтому разработка автоматизированной системы, которая рассматривает динамику изменений большого числа показателей, способна повысить эффективность процесса прогнозирования.

В данной работе задача раннего прогнозирования сепсиса рассматривается как проблема классификации многомерных временных рядов. Цель состоит в том, чтобы постоянно обновлять прогноз возникновения сепсиса у пациента на основе всех данных, имеющихся до этого момента [5].

Предполагается, что набор данных  $S$  состоит из  $N$  независимых пребывания пациентов в больнице. Для каждого пребывания  $i$ , имеется вектор ковариат обозначенный  $b_i \in R^B$ , такой как: вес, рост, возраст и т. д. В течение времени  $t_i = [t_{i1}, t_{i2}, \dots, t_{iT_i}]$  поступает информация о  $M$  разных показателях пациента, таких как: частота пульса,

температура, уровень глюкозы в крови, кровяное давление и т. д. Значения показателей обозначаются как  $Y_i = [y_{i1}, y_{i2}, \dots, y_{iM}] \in R^{T_i \times M}$ , где  $y_{im} \in R^{T_i}$  вектор снятых значений показателя  $m$  в течение всего пребывания  $i$ . На практике, в каждый момент времени  $t$  регистрируется только малое подмножество из всех показателей. Пребывания продолжаются разное время ( $T_i \neq T_{i'}$ ) и имеют свой уникальный набор времен наблюдения  $t_i$ .

Моделирование входящих параметров происходит с помощью процесса Гаусса. Условное распределение показателей пациента обновляется каждый раз при регистрировании новых данных. Это позволяет в каждый момент времени

предсказывать полный набор показателей пациента, сохраняя при этом долю неопределенности по каждому из них. Далее, условное распределение показателей пациента  $Z_i$  в момент времени  $x$  и  $i$ -ого пребывания вместе с вектором ковариат  $b$  поступают на вход классификатора, реализованного на основе РНС, где  $x$  – дельта времени обновления предсказывания. В ходе обучения РНС подаются метки о наличии или отсутствии болезни у пациента при текущем пребывании [6]. До того момента, как ошибка между работой классификатора и реальным значением не станет меньше  $k$ , происходит метод градиентного спуска, минимизирующий ошибку предсказывания (см. рисунок 1).

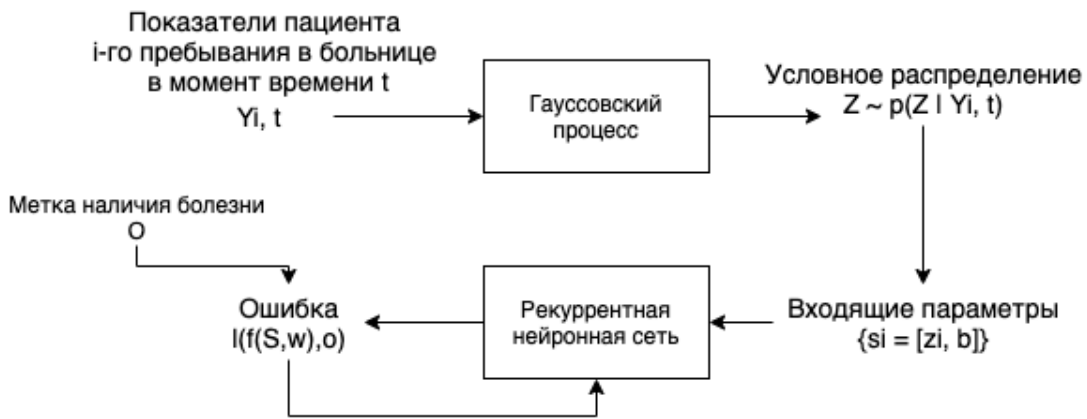


Рис. 1 – Модель системы.

Гауссовские процессы (ГП) способны моделировать нерегулярные временные ряды. Кроме того, они сохраняют неопределенность относительно дисперсии ряда в каждой точке, что важно в наших условиях, поскольку нерегулярность и отсутствие клинических временных рядов могут привести к высокой неопределенности для редко наблюдаемых переменных [7].

Ожидаемые значения показателей в предсказываемый момент времени рассчитываются на основе условного распределения обучающего набора  $u$ .

Пусть есть совместное распределение  $u$  и  $z$ :

$$\begin{pmatrix} y \\ z \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}, \begin{pmatrix} V_y & C_{yz} \\ C_{zy} & V_z \end{pmatrix}\right) \quad (1)$$

Тогда из уравнения (1) математическое ожидание и ковариационная матрица условного многомерного Гауссовского распределения принимает вид:

$$E(Z|Y = y) = \mu_z + C_{ZY}V_Y^{-1}(y - \mu_y) \quad (2)$$

$$E(Z|Y = y) = C_{ZY}V_Y^{-1}y = C_{ZY}(L_yL_y^T)^{-1}y = C_{ZY}L_y^{-T}L_y^{-1}y$$

где  $L_y$  положительно определенная матрица.  $L_y$  выводится с помощью разложения Холецкого:

$$V(Z|Y = y) = V_z - C_{ZY}V_Y^{-1}C_{YZ} \quad (3)$$

Где

$$y_i \sim N(0, \Sigma_i)$$

$$\Sigma_i = K^{T_i}$$

где  $K$  – корреляционная матрица, каждый элемент которой рассчитывается с помощью функции ядра  $k(t, t')$  [8]. В данной работе применяется рационально квадратичное ядро, так как оно полезно для моделирования зашумленных физиологических данных.

$$k(t, t') = \sigma^2 \left(1 + \frac{(t - t')^2}{2al^2}\right)^{-a}$$

где параметры  $a$  и  $l$  константы ядра.

Распределения рассматриваются как стандартно-нормальные. Поэтому  $\mu_z, \mu_y = 0$ .

Далее опишем алгоритм нахождения математического ожидания и ковариационной матрицы при известном обучающем наборе  $u$ :

$$V = LL^T$$

Вводим переменную  $m$ , такую что:

$$m = L_y^{-1}y \Rightarrow L_y m = y$$

После решения системы линейных уравнений получаем:

$$E(Z|Y = y) = C_{ZY}L_y^{-T}m$$

Вводим переменную  $s$ , такую что:

$$s = L_y^{-T}m \Rightarrow L_y^{-T}s = m$$

В итоге находим математическое ожидание

распределения показателей пациента при имеющихся данных  $y$ .

$$E(Z|Y = y) = C_{ZY}s \quad (4)$$

Аналогично находим ковариационную матрицу.

Имея математическое ожидание и матрицу ковариации из уравнений (2)–(4), можно построить условное распределение показателей на необходимом временном интервале.

$$z_{im}|y_{im} \sim N(E(Z|Y = y), V(Z|Y = y))$$

Пример априорного и условного распределений Гаусса показателя температуры одного из пациентов (см. рисунок 2).

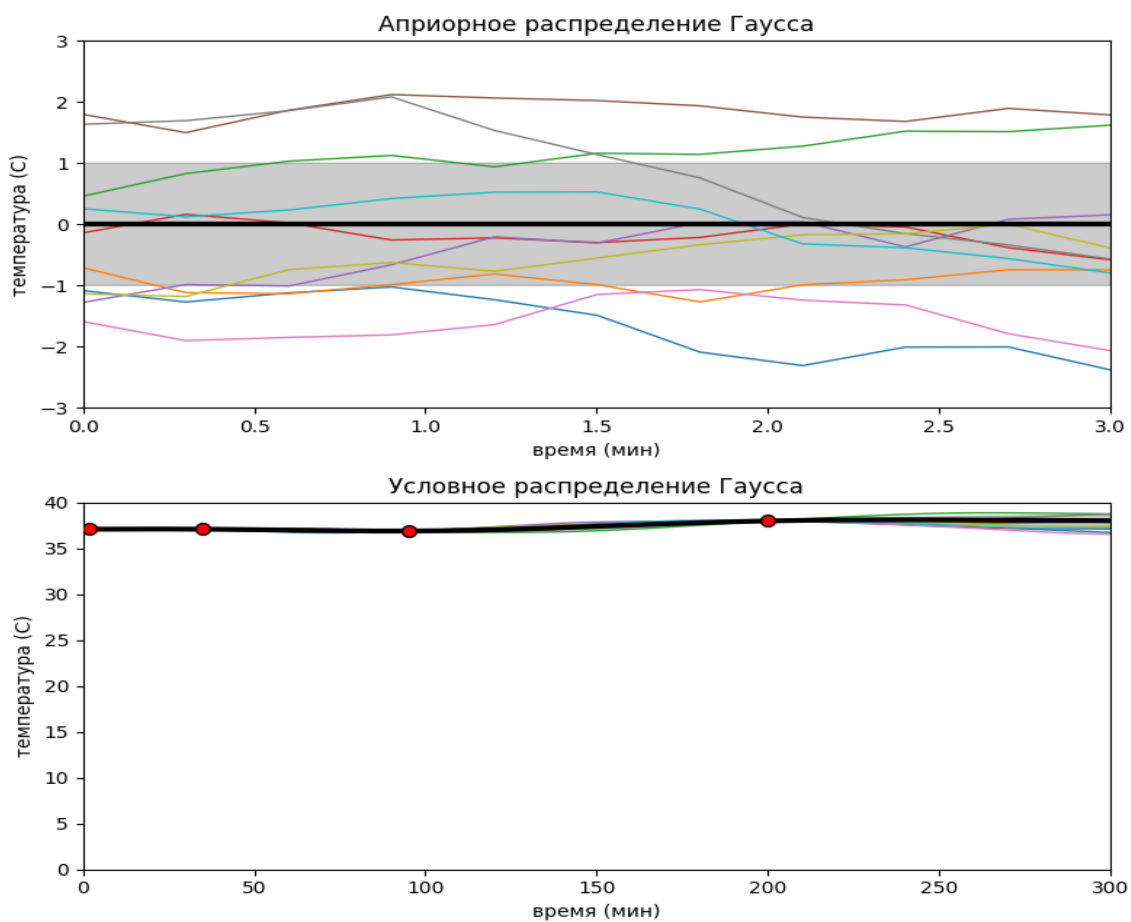


Рис. 2 – Априорное и условное Гауссовское распределения показателя температуры пациента.

На априорном распределении показаны 10 распределений Гаусса с нулевым математическим ожиданием и ковариационной матрицей равной значению ядра. После обновления априорного распределения построена условная вероятность на отрезке от 0 до 300 минут с учетом четырех снятий показателя на момент времени 230 минут от начала пребывания. Чем меньше частота снятия показателей, тем больше неопределенность.

Данные для исследования были взяты из медицинского информационного центра отделения интенсивной терапии (MIMIC) [9]. MIMIC –

огромная база данных процесса лечения пациентов, поступивших в медицинский центр Beth Israel Deaconess от 2001 до 2012 гг. Использование MIMIC базы данных для данного исследования подтверждено PhysioNet.

Из 23620 пребываний пациентов в ОИТ у 63 процентов имеется подозрение на инфекцию.

В рамках данной работы реализованы алгоритмы SIRS, SOFA и Гауссовский классификатор (ГК) с использованием РНС. Точность предсказания критерия SIRS за двенадцать часов до начала болезни равняется 70

процентам, за четыре часа – 78, за час 77. Точность критерия SOFA: 76, 84, 83 процентам и ГК: 81, 89, 91 процентам соответственно (см. рисунок 3).

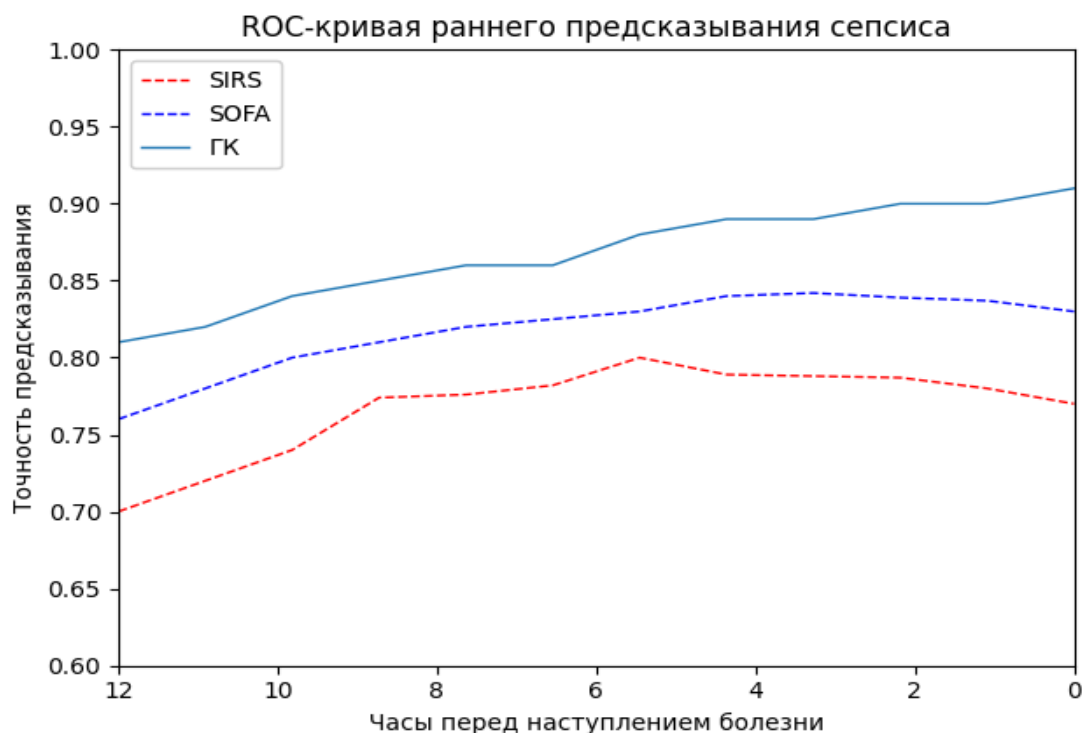


Рис. 3 – ROC-кривые раннего предсказания сепсиса систем ГК, SOFA и SIRS.

#### Заключение

Представленный подход для раннего выявления сепсиса классифицирует с высокой точностью многовариантные клинические временные ряды. Работа алгоритма показала более точные результаты по сравнению с общепринятыми критериями обнаружения сепсиса SIRS и SOFA.

Результаты исследования расширяют знания о системах прогнозирования возникновения сепсиса, и также способствует повышению эффективности процесса предсказания.

Конфликт интересов  
Не указан.

Conflict of Interest  
None declared.

#### Список литературы / References

1. Chakraborty RK and Burns B. Systemic Inflammatory Response Syndrome, 2019.
2. Lord JM, Midwinter MJ, Chen YF. The systemic immune response to trauma: an overview of pathophysiology and treatment, 2014.
3. Vincent JL, Moreno R, Takala J. Working Group on Sepsis-Related Problems of the European

Society of Intensive Care Medicine. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure, 1996.

4. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care, 2001.

5. Desautels T, Hoffman J, Barton C, Mao Q, Jay M, Calvert J, Das R. Pediatric severe sepsis prediction using machine learning, 2017.

6. Ryan J Delahanty, JoAnn Alvarez, Lisa M Flynn, Robert L Sherwin, and Spencer S Jones. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis, 2019.

7. Rasmussen, C. E. and Williams, C. K. I. Gaussian Processes for Machine Learning. The MIT Press, 2005.

8. Cheng-Xian Li, S. and Marlin, B. A scalable end-to-end gaussian process adapter for irregularly sampled time series classification. NIPS, 2016.

9. Johnson AE, Stone DJ, Celi LA, Pollard TJ. The MIMIC code repository: enabling reproducibility in critical care research, 2018.