Artificial intelligence-based model for automatic real-time and noninvasive estimation of blood potassium levels in pediatric patients

Hamid Mokhtari Torshizi1 , Negar Omidi2 , Mohammad Rafie Khorgami3 , Razieh Jamali4 , Mohsen Ahmadi5

¹Department of Biomedical Engineering and Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran, ²Department of Cardiology, Tehran Heart Center, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran, 3 Rajaie Heart Center and Department of Pediatric Cardiology, School of Medicine, Iran University of Medical Sciences, Tehran, Iran, ⁴Clinical Research Development Center, Mahdiyeh Educational Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran, ⁵Department of Biomedical Engineering, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution‑NonCommercial‑ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Torshizi HM, Omidi N, Khorgami MR, Jamali R, Ahmadi M. Artificial intelligence-based model for automatic real-time and noninvasive estimation of blood potassium levels in pediatric patients. Ann Pediatr Card 2024;17:116-23.

Address for correspondence: Dr. Mohsen Ahmadi, Department of Biomedical Engineering, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

E‑mail: dr.mohsen.ahmadi@gmail.com

Submitted: 22‑Mar‑2024 **Revised:** 26‑Apr‑2024 **Accepted:** 27‑Apr‑2024 **Published:** 20-Jul-2024

INTRODUCTION

As one of the main electrolytes, potassium plays a vital role in cellular membrane potential variations, especially in the heart.^[1] Normal cardiac function depends on regular sequential cardiac myocyte depolarization and repolarization. Any disruption in this circle may lead to cardiac conduction disorders and severe arrhythmia. The electrocardiograph (ECG) manifestations are usually related to potassium concentration, clinically measured as serum potassium levels.[2] Children with cardiac and kidney diseases are more susceptible to the effects of potassium changes.[3] Factors such as acute systemic illness, injection of potassium for electrolyte balance, and drugs and medicine intake can cause acute changes in potassium blood levels (K+).

It should be noted that 98% of K+ is intracellular (140 mEq/L) , and 2% is extracellular $(3.8-5.0)$ mEq/L).[4] Hypokalemia is the most common electrolyte imbalance in cardiac patients, delaying ventricular repolarization and downgrading conduction velocity, especially at the atrioventricular node. This can lead to various arrhythmias, such as sinus bradycardia and atrioventricular block.[5,6] Furthermore, hypokalemia can increase atrial and ventricular ectopic pulses and enhance digoxin's toxic effect.[7]

Until today, K+ levels have been measured in blood serum or plasma. Blood sampling in children is challenging, especially when frequent sampling is required. In some intensive care unit conditions, repeated blood sampling may result in anemia, unwanted cyanosis, or apnea in children with congenital heart diseases. Furthermore, this method is invasive, expensive, and requires blood samples and some time to get the test results. On the other hand, K+ measurements with laboratory technical processes require clotting preparation, which could take some time. This can yield abnormal potassium levels due to time-lapse.[8-10]

In the face of these challenges, we recently designed a method for serum potassium concentration quantification from ECG analysis. Although a few studies have been done using machine learning algorithms, $[11-13]$ they must be refined to be more practical. In this study, we validated our K + estimator and tested it on a large group of patients. Potassium value extraction using a single lead would permit its use in wearable, wireless ECG patches and possibly in implantable loop recorders and cardiac implantable electronic devices (pacemakers and defibrillators).

METHODS

Study subjects

A total of 428 hospitalized patients in the Rajaie Cardiology and Medical Research Center and Tehran Heart Center Emergency Department were recruited from December 2021 to June 2022. The ECG features of patients were evaluated. The patient's serum K+ level and ECG were taken within 2 h of admission. Patients having a history of heart failure, end-stage renal disease (ESRD), bundle branch block, strain pattern in ECG, premature ventricular contraction, and digoxin use were excluded from the study. A specialist nurse took the ECGs. Patients whose ECGs had noise or artifacts for any reason were excluded from the study. Information about the eligible patients, including their basic demographic information, has been collected through interviews and questionnaires. The patients or their representatives provided informed consent to use their data for the study. The Institutional Ethical Committee approved the study (Ethical Code: IR.SBMU.MSP.REC.1398.982).

Feature extraction

We recorded each patient's ECG over 2.8 s using a Philips IntelliSpace ECG (12 channels). The data were transferred from the device's memory to an external memory to analyze the ECGs. We then developed a program in Python language on the Windows platform that can extract the amplitude and time values of P, Q, R, S, and T from lead II. To remove noise and unwanted values, we calculated the median QT intervals and excluded values that differed more than 20 ms from the median. In total, 16 features were computed using these points, as shown in Figure 1.

Preprocessing

We used z-score standardization according to the following formula:

$$
Z_i = \frac{X_i - X}{S}
$$

- X_i is a data point $(X_1, X_2 \ldots X_n)$
- \bar{X} is the sample mean
- S is the sample standard deviation (SD).

Dimensional reduction

In each data set, there is a possibility of having two features that are highly correlated. Two identical

Figure 1: Display of features extracted from the cardiac signal

features naturally create additional redundancy. The cross-correlation matrix was used for dimensionality reduction. For every two signal features with a correlation >0.7, one was removed [Figure 2].

Machine learning

Analysis foundation

The programming language used for signal processing, machine learning, and survival analysis was Python software (Version 3.10, released in October 2021). The integrated development environment used for analysis was RStudio (1.4.1106, RStudio Team [2020]. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.), an integrated development environment for Visual Studio.

Regression machine learning algorithms

A regression model is one of the most common types of supervised learning in machine learning. We encompassed linear and nonlinear algorithms, including support vector regression (SVR), decision tree, random forest, and polynomial regression. We used Pearson's correlation coefficient to evaluate the linear relationship between the ECG characteristics and the serum potassium levels. We assessed the nonlinear relationship using decision trees and random forest algorithms.

Linear regression

In the simplest definition, linear regression tries to match many data points with a straight line in two-dimensional space or a plane in three-dimensional space. This type of regression examines the linearity between observations and targets and tries to show the relationship between them as a linear equation or a weighted sum function. If a data set has n features x, the target y can be defined as follows:

$$
y = W_0 + W_1 X_1 + W_2 X_2 + \dots + W_n X_n = W^T X
$$

The linear regression model, specifically the values of w, is adjusted and calculated based on the training data. The weighted values are trained to minimize the mean square error (MSE) (the mean square difference between reality and prediction). If we have m training samples, the cost function J (w) can be expressed according to the following formula:

$$
J(w) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} (\hat{y}(x^{i}) - y^{i})^{2}
$$

which $\hat{y}(x) = w^T x^i$ is the prediction.

The value of w should be optimized so that the value of the cost function is minimized. For this, the gradient reduction method is used. The first-order derivative is obtained as follows:

$$
\Delta W = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} (\hat{y}(X^i) - y^i) X^i
$$

The weight vector, w, can be updated by defining the learning rate, η, and the gradient defined:

$$
W := W + \eta \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} (\hat{y}(x^{i}) - y^{i}) x^{i}
$$

The weights of this algorithm are updated and trained several times depending on the value of η. The prediction of a new value can be calculated using the following formula:

$$
y' = w^T x'
$$

Polynomial regression

In polynomial regression, the relationship between the independent variable x and the dependent variable y is modeled as nth-degree polynomial in x.

Figure 2: Cross-correlation matrix. One of the two variables with a correlation >0.7 is a candidate for elimination

Regression tree algorithm

In this model, the algorithm works by recursively partitioning the data into subsets based on the predictor variables and then fitting a simple model for each subset. This procedure continues until a stopping criterion is met, such as a maximum depth or a minimum number of data points in a subset.

Algorithms such as 3, C4.5, Classification and Regression Trees (CART), and Chi-square automatic interaction detection have been introduced to form a tree with a suitable structure. Currently, the CART algorithm is considered the most important decision tree algorithm. In this algorithm, a tree is built based on binary division. Each node splits into left and right branches. In each section, it greedily searches for the most important combination of one feature and its value. All possible combinations are evaluated using a mathematical function. However, in this algorithm, using the best-selected feature and its value as the dividing point, the data set is divided as follows:

- Samples with features equal to or greater than the value of the dividing point of the right branch
- Remaining examples from the left branch.

The process of regression using the decision tree is similar to its classification mode, but due to its continuous nature, it has two differences: (1) the correctness of the decision point is measured by the MSE of the two branches. The mean squared error of the branch can be considered the variance of all target values. The smaller the MSE, the better the division. (2) The average value of the target at the end node is converted into a leaf value instead of the majority label.

The stopping criterion can be the number of leaf divisions, the minimum MSE, or a combination of both.

Random forest

One of the weaknesses of the decision tree method is high variance. This weakness can be reduced to a large extent by combining decision trees. Random forest is a group learning method that involves combining several trees. Each tree is trained, and several random features are sampled at each tree node. The average of the regression results from all the decision trees is assigned to the final decision.

Estimation with support vector regression

The SVR is part of the support vector family. In SVR, the goal is to find a super plane with slope w and bias b so that two super planes with specifications wx + b = ε and wx + $b = -\epsilon$ cover most of the desired data.

Theoretically, the ideal superplane is as flat as possible. According to the figure, most data are located in the ε bands of the desired superplane. The optimal extraction of w and b depends on the implementation of the following two conditions: 2–6 model validation:

1. Minimize lwl

2.
$$
\left| y^{i} - \left(wx^{i} + b \right) \right| \leq \varepsilon
$$

In this study, we used 20% of the population to evaluate our model performance, while 80% was used to train the model. The estimated potassium level was calculated from the obtained ECG data using the corresponding patient‑specific potassium prediction model developed during the training phase. To assess the accuracy, we calculated the mean absolute error and the mean absolute value of the difference between the estimated and measured potassium levels for each patient.

Variable importance

We used the random forest algorithm capability to value the variables.^[14]

RESULTS

Among the 463 patients admitted to the hospitals in 1 year, 35 patients were excluded from the study due to high noise and distortion of the ECG. Among the study population, 56% were boys and were aged between 1 and 14 years (mean \pm SD: 5 \pm 3).

The potassium chart of these patients is shown in Figure 3. The results of the cross-correlation matrix of data are shown in Figure 2.

Based on the information in Figure 2, we remove one of the two parameters that correlate more than 0.7. The features used to train the regression methods (algorithms) were PR, Ps, PT, Twidth, QS, QR, QT, RS, RT, ST, and RtoT. Table 1 shows the efficiency of each regression algorithm based on the MSE. As indicated in the table, the polynomial method has the lowest accuracy, and the random forest method has the highest measurement accuracy.

Considering that most of our studied patients have a potassium level between 4 and 4.5, we used the scatter diagram to get better feedback than regression methods. Figure 4 shows the scatter diagram for different approaches.

In Figure 5, the importance of each feature by the random forest algorithm is shown in percentage terms. Figure 6 shows how the decision tree algorithm yields decisions.

DISCUSSION

Acute electrolyte disturbance, especially hyperkalemia in children, is life-threatening and requires prompt attention. In addition to correctly determining the K+ level, the on-time result is critical because many children with acute illness need immediate medical attention. Blood sampling is a well-known, reliable method for evaluating this electrolyte. Evaluation of the K+ level noninvasively

has always been desirable, and attempts have been made to determine the K+ level with high accuracy and sensitivity.

Figure 3: Histogram of blood potassium serum level

The effect of potassium on the ECG cardiac signal has been known for many years.[15-17] So far, several studies, generally based on T-wave morphology, have been conducted to determine serum potassium levels.[18-22] ECG markers, defined as a specific time interval or range of neural signals, are noise-prone.^[23] To solve this problem, many studies have investigated T-wave morphology over a long period.[24-26] Significant challenges with these studies have

Table 1: Efficiency of regression methods based on mean squared error

MSE: Mean square error, SVR: Support vector regression

Figure 4: Relation between predicted potassium serum level by machine learning model and measured potassium serum level via blood samples. Scatter diagram of (a) Linear regression, (b) Polynomial regression, (c) Support vector regression, (d) Regression tree, (e) Random forest

been the definition of morphologies based on complex mathematical rules, the long time to calculate optimal parameters (several hours), and the low sample size.

The purpose of this study is to design and validate an online and noninvasive potassium level extraction

Figure 5: Features' importance by the random forest algorithm. The influence of a variable is shown as a percentage

technique based on a machine learning algorithm. We allocated 2.8 s to determine the serum level of potassium. This time interval ensures that several cardiac cycles are considered. Another strength of this study is the application of filtering to remove the effect of unwanted distortion on ECG parameters. Therefore, if a part of the signal is distorted, the result of evaluating the potassium level is still reliable. Another advantage of this method is that only one lead is considered. This helps to commercialize the mentioned method and assess the potassium level remotely. Considering the large sample size, it can be assumed that all T-wave morphologies were covered. We are developing our algorithm in such a way that we can analyze different types of ECG signals.

So far, attempts have been made to measure blood potassium levels using cardiac signals. Aslam et al. studied 74 ESRD patients in 2002.[27] In their study, they tried to provide a linear relationship between T-wave amplitude or T-wave to R-wave ratio and the serum potassium level in a patient's blood. Similarly, Szerlip et al. [28] could not provide a relationship between the T-wave to R-wave ratio and the serum potassium level in individuals' blood.

Figure 6: How to allocate blood serum potassium level based on the input characteristics in the decision tree algorithm

Torshizi, et al.: Automatic real-time estimation of blood potassium level

In 2020, with the introduction of two parameters, T-right slope and T-amp, a group of researchers tried to discover the potassium serum level with ECG.[29] By examining two parameters (T-amp and T-right slope) on five lead waveforms (V3, V4, V5, V6, and II), they found that T-wave-based features were not correlated with serum potassium level. Yasin et al. investigated the potassium serum level and ECG of 21 dialysis patients.[30] The average absolute error between estimated and blood potassium was 0.38–0.32 mEq/L. The sample size of our study and the wide range of serum potassium levels are the strengths of our work compared to theirs.

Another study tried to quantify the amount of potassium using the cardiac signal and the serum potassium level in the blood of 12 patients.[22] They used the following variables for the regression operator in their work: the slope of the T-wave downstroke (T right slope), the amplitude of the T-wave (T-amplitude), the center of gravity (COG) of the T-wave (T COG), the ratio of the amplitude of the T-wave to the amplitude of the R-wave (T/R amplitude), and the COG of the last 25% of the area under the T-wave curve (T4 COG). They have recommended using the cardiac signal as a noninvasive method. However, due to the small sample size of their study, more research is required. In our study, we tried to overcome the weaknesses of previous studies by considering a large sample size with a wide range of serum potassium levels and more cardiac signal features and applying machine learning methods.

The sample size of this study has made it possible to examine all the clinical factors of the cardiac signal. Based on the results in Figures 5 and 6, the PT parameter is considered the most important influencing factor in predicting serum potassium levels. This means the combination of PQ, QRS, and ST intervals is influential. However, the second most important parameter in determining serum potassium levels is the difference in R- and T-wave amplitudes. This finding is similar to studies that have considered the slope of the T-wave as an important parameter in the determination of serum potassium.

As shown in Figure 4, the linear and polynomial methods used in most studies do not have adequate predictive power. Although the polynomial method has high accuracy in the training phase, it has the lowest accuracy in the test phase [Table 1]. The phenomenon of overfitting causes this. The linear method also predicts all values in the 4–4.5 mmol range. Although the decision tree method works well for wide potassium levels, it is unsuitable for low potassium (under 4 mmol). Prediction of potassium level through the SVR method only works somewhat well in medium values. However, the random forest method has solved the problem of the decision tree algorithm to a great extent and has improved the detection of potassium in low amounts. Finally, it can be said that in this study, the random forest algorithm is more efficient than other algorithms.

Besides potassium, other factors such as other electrolytes and the location of the leads affect the cardiac signal. Despite all the sources influencing the cardiac signal, we calculated the serum potassium level with an average error of 0.3 in this study. Perhaps, in future studies, the effect of other electrolytes on ECG can be processed, and by considering their effects on the cardiac signal, an increase in the accuracy of predicting blood potassium can be attained.

CONCLUSIONS

We defined a comprehensive noninvasive method for evaluating K+ level in pediatrics based on ECG signal. This study may evolve a noninvasive portable method for determining K+ level by monitoring ECG signal results and enabling quick measurement of K+ level changes in response to systemic conditions in acutely ill children and adults.

Acknowledgment

The present article has been extracted from the PhD thesis written by Hamid Mokhtari Torshizi and was financially supported (Grant No, 27067) by the School of Medicine, Shahid Beheshti University of Medical Sciences (Registration No, 22201). We thank the staff of Shahid Rajaei Hospital for their efforts.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- 1. Joukar S. A comparative review on heart ion channels, action potentials and electrocardiogram in rodents and human: Extrapolation of experimental insights to clinic. Lab Anim Res 2021;37:25.
- 2. Raffee LA, Alawneh KZ, Ababneh MJ, Hijazi HH, Al Abdi RM, Aboozour MM, et al. Clinical and electrocardiogram presentations of patients with high serum potassium concentrations within emergency settings: A prospective study. Int J Emerg Med 2022;15:23.
- 3. Wieërs ML, Mulder J, Rotmans JI, Hoorn EJ. Potassium and the kidney: A reciprocal relationship with clinical relevance. Pediatr Nephrol 2022;37:2245-54.
- 4. Ferreira JP, Butler J, Rossignol P, Pitt B, Anker SD, Kosiborod M, et al. Abnormalities of potassium in heart failure: JACC state-of-the-art review. J Am Coll Cardiol 2020;75:2836-50.
- 5. Michaud GF, Sticherling C, Tada H, Oral H, Pelosi F Jr., Knight BP, et al. Relationship between serum potassium concentration and risk of recurrent ventricular

tachycardia or ventricular fibrillation. J Cardiovasc Electrophysiol 2001;12:1109-12.

- 6. Fisch C, Knoebel SB, Feigenbaum H, Greenspan K. Potassium and the monophasic action potential, electrocardiogram, conduction and arrhythmias. Prog Cardiovasc Dis 1966;8:387-418.
- 7. Aghajani M, Imani A, Kardar G, Faghihi M, Aghajani M. Unfavorable left ventricular remodeling due to experience of chronic sleep restriction after myocardial infarction: The role of matrix metalloproteinases & oxidative stress. Cardiovasc Pathol 2023;62:107460.
- 8. Meng QH, Wagar EA. Pseudohyperkalemia: A new twist on an old phenomenon. Crit Rev Clin Lab Sci 2015;52:45-55.
- 9. Cooper LB, Savarese G, Carrero JJ, Szabo B, Jernberg T, Jonsson Å, et al. Clinical and research implications of serum versus plasma potassium measurements. Eur J Heart Fail 2019;21:536-7.
- 10. Hawkins R. Measurement of whole-blood potassium Is it clinically safe? Clin Chem 2003;49:2105-6.
- 11. Salari N, Hosseinian Far A, Mohammadi M, Ghasemi H, Khazaie H, Daneshkhah A, et al. Detection of sleep apnea using machine learning algorithms based on ECG Signals: A comprehensive systematic review. Expert Syst Appl 2022;187:115950.
- 12. Chakraborty A, Chatterjee S, Majumder K, Shaw RN, Ghosh A, editors. A comparative study of myocardial infarction detection from ECG data using machine learning. In: Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021. Lecture Notes in Networks and Systems: Springer; 218;2022.
- 13. Udawat AS, Singh P. An automated detection of atrial fibrillation from single-lead ECG using HRV features and machine learning. J Electrocardiol 2022;75:70-81.
- 14. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics 2008;9:307.
- 15. Lanari A, Chait LO, Capurro C. Electrocardiographic effects of potassium. I. Perfusion through the coronary bed. Am Heart J 1964;67:357-63.
- 16. Laks MM, Elek SR. The effect of potassium on the electrocardiogram: Clinical and transmembrane correlations. Dis Chest 1967;51:573-86.
- 17. El Sherif N, Turitto G. Electrolyte disorders and arrhythmogenesis. Cardiol J 2011;18:233-45.
- 18. Attia ZI, DeSimone CV, Dillon JJ, Sapir Y, Somers VK, Dugan JL, et al. Novel bloodless potassium determination using a signal-processed single-lead ECG. J Am Heart Assoc 2016;5:e002746.
- 19. Severi S, Corsi C, Haigney M, DeBie J, Mortara D, editors. Noninvasive Potassium Measurements from ECG Analysis during Hemodialysis Sessions. 2009 36th Annual Computers in Cardiology Conference (CinC). IEEE; 2009.
- 20. Corsi C, DeBie J, Napolitano C, Priori S, Mortara D, Severi S, editors. Validation of a Novel Method for Non‑Invasive Blood Potassium Quantification from the ECG. 2012 Computing in Cardiology. IEEE; 2012.
- 21. Corsi C, Cortesi M, Callisesi G, De Bie J, Napolitano C, Santoro A, et al. Noninvasive quantification of blood potassium concentration from ECG in hemodialysis patients. Sci Rep 2017;7:42492.
- 22. Dillon JJ, DeSimone CV, Sapir Y, Somers VK, Dugan JL, Bruce CJ, et al. Noninvasive potassium determination using a mathematically processed ECG: Proof of concept for a novel "blood-less, blood test". J Electrocardiol 2015;48:12-8.
- 23. Palmieri F, Gomis P, Ferreira D, Ruiz JE, Bergasa B, Martín Yebra A, et al. Monitoring blood potassium concentration in hemodialysis patients by quantifying T-wave morphology dynamics. Sci Rep 2021;11:3883.
- 24. Bukhari HA, Sánchez C, Srinivasan S, Palmieri F, Potse M, Laguna P, et al. Estimation of potassium levels in hemodialysis patients by T wave nonlinear dynamics and morphology markers. Comput Biol Med 2022;143:105304.
- 25. Palmieri F, Gomis P, Ruiz JE, Ferreira D, Martín-Yebra A, Pueyo E, et al. ECG-based monitoring of blood potassium concentration: Periodic versus principal component as lead transformation for biomarker robustness. Biomed Signal Processing Control 2021;68:102719.
- 26. Rodrigues AS, Petrėnas A, Paliakaitė B, Kušleikaitė Pere N, Jaruševičius G, Bumblytė IA, et al. Noninvasive monitoring of potassium fluctuations during the long interdialytic interval. IEEE Access 2020;8:188488-502.
- 27. Aslam S, Friedman EA, Ifudu O. Electrocardiography is unreliable in detecting potentially lethal hyperkalaemia in haemodialysis patients. Nephrol Dial Transplant 2002;17:1639-42.
- 28. Szerlip HM, Weiss J, Singer I. Profound hyperkalemia without electrocardiographic manifestations. Am J Kidney Dis 1986;7:461-5.
- 29. Chen H, Maharatna K. An automatic R and T peak detection method based on the combination of hierarchical clustering and discrete wavelet transform. IEEE J Biomed Health Inform 2020;24:2825-32.
- 30. Yasin OZ, Attia Z, Dillon JJ, DeSimone CV, Sapir Y, Dugan J, et al. Noninvasive blood potassium measurement using signal-processed, single-lead ECG acquired from a handheld smartphone. J Electrocardiol 2017;50:620-5.