Development of a patient health monitoring system based on the internet of things with a module for predicting vital signs

Yerlan Zaitin[1], Madina Mansurova[1], Murat Kunelbayev[2], Gulnur Tyulepberdinova[1], Talshyn Sarsembayeva[1], Adai Shomanov[3]
[1]Department of Arfificial Intelligence and Big Data, Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Institute of Information and Computer Technologies, Almaty, Kazakhstan
[3]Nazarbayev University, Astana, Kazakhstan

Article Info

Article history:

Received Oct 13, 2023
Revised Nov 1, 2023
Accepted Nov 2, 2023

Keywords:

Ambient assissted living
Disease prediction
Health monitoring
Internet of things
Machine learning

ABSTRACT

Recent issues related to human health in the world have shown the importance of telemedicine considering necessities to perform the remote monitoring of patients. In this study, using a patient smart monitoring system (PSMS), we collected 5,000 samples of heart rate and blood saturation vital signs from 4 volunteers and tried to find better correlation algorithms to develop a module to predict what these vital signs will be in the next 60 seconds. The following regression algorithms recurrent neural network (long short-term memory) (RNN(LSTM)), autorregresive integrated moving average (ARIMA), value-added reseller vector autoregression (VAR) were used to forecast the patient's state of health in the next 60 seconds. Further, the support vector machine (SVM) and Naive Bayes classification algorithms use the data forecasted by the regression algorithms as input to predict the health status of the patients. When comparing algorithms, we focused on the F measure, a metric used to evaluate the performance of machine learning algorithms. As a result, RNN(LSTM) and SVM showed the performance score value of machine learning algorithms F 0.84, RNN(LSTM) and Naive Bayes 0.81, VAR and SVM 0.82, and VAR and Naive Bayes 0.75. Compared to them, the correlation of ARIMA regression algorithms and SVM classification showed a better F score of 0.86 for machine learning algorithms than the others.

Corresponding Author:

Yerlan Zaitin
Department of Arfificial Intelligence and Big Data, Faculty of Information Technology
Al-Farabi Kazakh National University
71 al-Farabi Ave., Almaty, Kazakhstan
Email: yzaitin@gmail.com

1. INTRODUCTION

One of the latest advances in digital health is telemedicine [1]. With the help of mHealth and internet of things (IoT) technologies, patients are monitored remotely at home [2]. Many diseases require long-term monitoring of the patient during treatment, such as chronic diseases, and heart diseases. In such cases, the IoT device must be able to effectively perform real-time monitoring [3]. In addition, the module, which predicts the physiological data of patients using machine learning algorithms, reduces the high number of deaths due to diseases such as cardiovascular diseases and chronic respiratory diseases [4]. Heart rate and blood oxygen saturation are the main fundamental data when assessing the condition of such patients [5].

Until now, several technologies have been used to remotely monitor these two patient parameters and predict their values. Oscillometer sensors are the most commonly used type of sensor for remote monitoring of blood pressure (BP). These sensors use the oscillations caused by the pulsations of blood flow to measure systolic and diastolic BP [6]. Oscillometer sensors are non-invasive and are typically worn on the wrist or upper arm. They are easy to use and provide accurate readings, making them a popular choice for remote monitoring of BP in patients with cardiovascular disease (CVD). Plethysmography sensors used for remote monitoring of blood pressure. These sensors use light to measure the volume changes in blood vessels, which are caused by the pulsations of blood flow. Plethysmography sensors are typically worn on the finger or earlobe and are considered to be more accurate than oscillometer sensors [7]. However, they are expensive and may be less convenient for patients to use.

The measurement of SpO2 is important as it provides insight into the oxygenation status of the body, which is critical for the management of CVD. Pulse oximetry is the most commonly used method for measuring SpO2 in patients with CVD. Pulse oximetry sensors consist of a light-emitting diode (LED) and a photodetector, which are placed on the patient's finger or earlobe [8]. The LED emits light of different wavelengths, and the photodetector measures the amount of light absorbed by the patient's tissue. The sensor then calculates the SpO2 based on the ab-sorption of light at different wavelengths. A study by Xu *et al.* [9] compared the accuracy of pulse oximetry sensors placed on the finger and earlobe in patients with CVD. The study found that the sensors placed on the earlobe were more accurate than those placed on the finger. The authors also found that the accuracy of the sensors was not affected by the patient's age or gender.

The literature details several machine learning algorithms that have been used to predict future values of the above physiological data. Recurrent neural network (RNN) with long short-term memory (LSTM) is a type of neural network that can model the time dependence of subsequent data. They widely apply time series forecasting, including market investment forecasting and energy consumption forecasting [10]. LSTM networks also detect anomalies in time series data [11]. However, the limitation of LSTM networks is their high computational cost and involves large amounts of data for training. Autorregresive integrated moving average (ARIMA) is a classic statistical approach that models a time series as a combination of its past values, differences, and lagging forecast errors. ARIMA is widely used in various applications such as electricity load forecasting and traffic flow forecasting [12]. However, ARIMA assumes that the data is stationary, which may not always be the case in real scenarios. Also, ARIMA may not be well suited for long-term forecasting. Vector autoregression (VAR) is a statistical model that generalizes ARIMA to multiple time series variables. It models relationships between variables and uses the lagging values of each variable as predictors. VAR has been used to predict macro-economic variables [13]. However, like ARIMA, VAR assumes that the data is stationary and can be poorly used for long-term forecasting.

According to our research, the disadvantage of previous studies is the lack of an integrated system to monitor and predict patients' physiological data in real time. In addition, finding the best correlation of regression and classification algorithms for predicting the next value of physiological parameters and dividing these predicted values into classes depending on the health status of patients is still an unsolved problem. In this work, IoT hardware and software technologies were combined to create a complete system to collect important physiological data of patients-heart rate and blood saturation using a unique device and predict the future value of these data. To the best of our knowledge, this is the first system that can monitor patients' physiological data in real time and predict their future value.

The purpose of this work is to improve the quality of patient care by creating an integrated system for predicting and monitoring vital signs. By presenting patients' potential health trajectories in a timely manner, this system eases pressure on acute care facilities and strengthens disease prevention strategies. In the following sections of this article, we aim to provide a comprehensive review of our methods and results.

2. METHOD

In this work, we used the technique of remote patient monitoring as shown in Figure 1. The monitoring system's data flow diagram (DFD) is displayed in Figure 2. Patient smart monitoring system (PSMS) is based on two parties communicating with one another. There are two more databases at DFD. One is for patient and health data, and the other is for information on the doctor. The connection and data processing are the main goals of the PSMS process. The two main data flows in DFD are from entities to databases. They are patient health information and patient/doctor information. The PSMS process uses "PatientDB" to collect data on patient health (heart rate and saturation) and information (full name, phone number, address, e-mail), as well as doctor information (full name, position, and address). The PSMS process is divided into "AI monitoring" and "Data processing," as well as another entity known as "Ambulance service." The addition of a new entity is what led to the emergence of three data flows, called "Call" and "Notification" to inform "Ambulance service" and "Help" entities led to demonstrate the functioning of an ambulance, respectively. PSMS is a "doctor" and a "AI monitoring" system that uses data that has been processed and warned as feedback from a doctor or as basic health data with some recommendations. Figure 3 shows the materials used in the device for collecting vital signs data of patients. A device was created based on ESP32 and a mobile application using Flutter framework in Dart language. Figure 4 shows the mobile application prototypes. Since the ESP32 has a built-in Bluetooth module, communication with a mobile application written in Dart using the Flutter framework was implemented. If there is no Wi-Fi connection, the app will be offline and will only show data from your phone. When connected to Wi-Fi, the app is online, thus sending data to the firebase server. Thus, we collected such vital signs data using heart beat and SpO2 sensors. For experiments based on disease prediction system, we obtain vital signs data from PSMS. The dataset contains heart rate and SpO2 vitals for 4 different volunteer-based case studies (the dataset is pre-filtered to check for any outliers (i.e., low heart rate (HR), low SpO2, and high HR) processed). The proposed model aims to predict vital signs sixty seconds into the future. The vital signs prediction system is implemented using separate regression and classification models (i.e., to predict vital signs within the next 60 seconds) and different ranges or levels: low, normal and high to indicate the patient's condition. For classification, we use a combination of different ranges (or levels) of vital signs to assign a unique marker or class of patient status based on the advice and recommendation of a healthcare professionals. The prediction model first predicts a vital sign for a specific time period (i.e., 60 seconds). Additionally, these predicted vital signs are fed into a classification model (a machine learning-based model) that classifies the patient's condition into one of the 7 classes. The class sign provides the general condition of the patient and helps cardiovascular disease to know the cause of the patient's condition (vital sign). This early detection helps explain abnormal vital signs (i.e., low HR, high HR, and low SpO2).

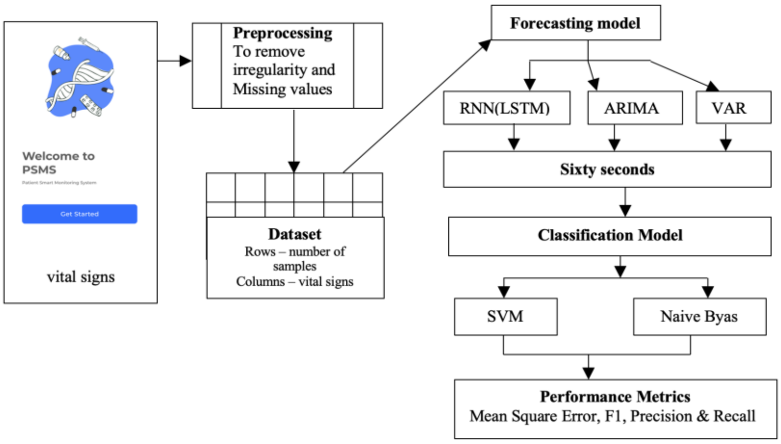

Figure 1. Proposed methodology for predicting vital signs

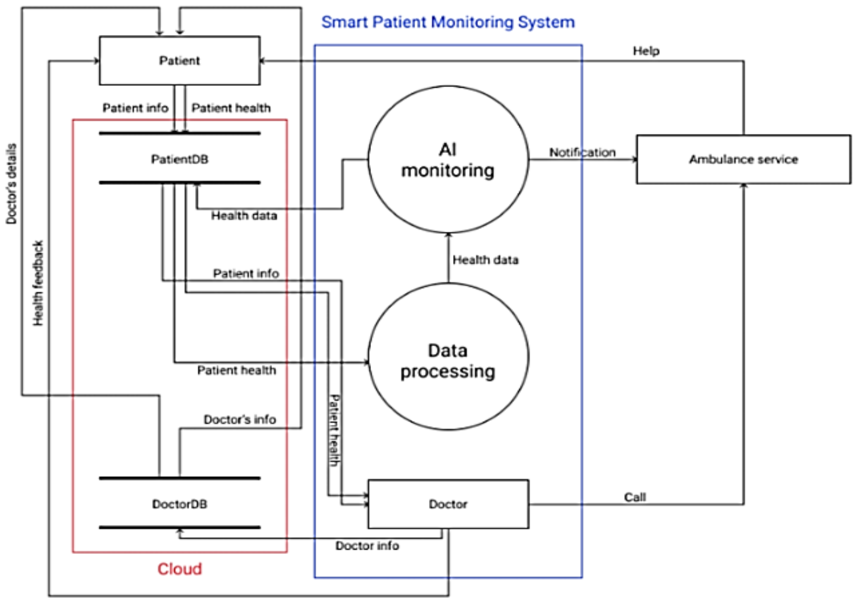

Figure 2. Data flow diagram of the PSMS

We divide the dataset into two parts to train and test the machine learning model. To train the regression models (to forecast vital signs), we used 5,000 samples of vital signs at 10 millisecond intervals. These sampling units are based on 10-minute (50,000 samples at 10-millisecond intervals) patient monitoring. For a critically ill patient, the first 60 seconds are usually critical and require immediate assistance from caregivers or medical experts. That's why we focus on predicting vital signs within the next 60 seconds to provide timely patient care. To predict 60-second vital signs, we first performed training of a machine learning predictor using 50,000 samples. For evaluation purposes, several benchmarks are used to evaluate the performance of each algorithm. These standard measures include F-measure, recall, precision, mean square error (MSE). After applying these metrics, the results are explained in detail in chapter 4.

Malasinghe's research used correlation of binomial, trinomial, quadratic polynomial regression, and Naive Bayes, support vector machine (SVM) classification algorithms and the University of Queensland vital signs dataset [14]. Their study gave the best F value of 0.83 due to the correlation between quadratic polynomial and SVM algorithms. The scientific novelty of our study is that we collected vital signs data using our own PSMS device and used RNN(LSTM), ARIMA, VAR regression and Naive Bayes and SVM classification algorithms. As a result, the best F-score of 0.86 was achieved due to the correlation of ARIMA and SVM algorithms of this indicator.

This work has scientific and theoretical significance due to a thorough study of the mathematical foundations of algorithms for RNN(LSTM), ARIMA and VAR. They were used to create a predictive module for predicting heart rate and blood saturation for the next 60 seconds. The practical significance arises from the translation of this theoretical framework into a real application using Python, which allows you to carefully compare these algorithms and determine the most appropriate one. Moreover, the integration of these predicted results with SVMs and Naive Bayes algorithms for classification highlights the practical usefulness of the study, offering a universal set of tools for both accurate regression and informative classification of vital signs, thereby speeding up the medical decision-making process. and improving patient care.

A clever combination of cutting-edge algorithms, such as RNNs(LSTM), ARIMA, and VAR, is used in this study to forecast vital signs within the next 60 seconds. This method provides a thorough grasp of the data's temporal dynamics. On the basis of this, the study employs SVM and Naive Bayes algorithms, employing the projected values as input, boosting the prediction capabilities for classification tasks. The study's unique contribution is the discovery of correlations between regression and classification methods, which represents substantial advancement. Regression and classification are linked in this interconnected approach, demonstrating a novel methodology that offers more accurate analysis of medical data and makes room for broader applications.

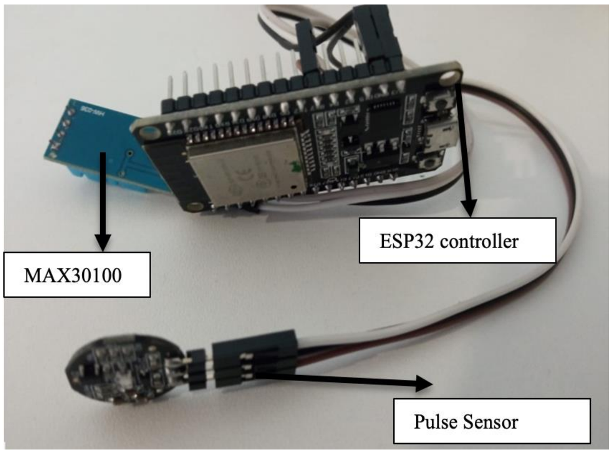

Figure 3. Devices used in PSMS

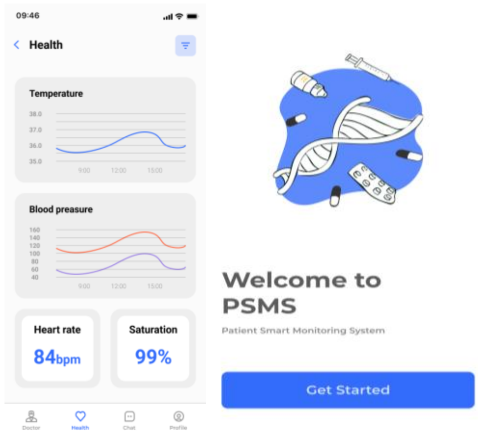

Figure 4. Mobile application prototypes

2.1. Data collection

The first step is to collect the vital sign data. We used the data obtained by the PSMS device. The dataset contains heart rate and blood oxygen vital signs. The dataset contains 5 hours data of different volunteers. The details about dataset attribute and patients records in millisecond with 10-millisecond interval are mentioned in Table 1.

Table 1. List of vital signs and volunteers' records

Index	Vital signs	Volunteers file name	Number of samples (with 10 ms interval)
1	Time	Case01	5465
2	Relative time (MS)	Case02	5222
3	HR	Case03	5252
4	SpO2	Case04	5850

2.2. Preprocessing and normalization

Data normalization is associated with the removal of redundant (duplicate) and partial data. we merge different patient files into one file. After that, we delete irrelevant vital signs from these files (it can be seen in Table 2). The dataset contain 5,789 samples for all 4 patients. After pre-processing, we apply normalization of the relevant vital signs. Remove all unnecessary (duplicate) data and the missing values as shown in Table 3. After preprocessing up to 10% of data is removed.

Table 2. Sample raw data of a volunteer

Index	Vital signs	Volunteers file name	Number of samples (with 10 ms interval)
1	Time	Case01	5,465
2	Relative time (MS)	Case02	5,222
3	HR	Case03	5,252
4	SpO2	Case04	5,850

Table 3. Sample preprocessed and normalized data

Time	Relative time (MS)	HR (BPM)	SpO2(%)
00:00:00_000	0	78	98
00:00:00_010	10	85	94
00:00:00_020	20	79	82
00:00:00_030	30	89	83
00:00:00_040	40	86	97
00:00:00_050	50	85	78

2.3. Vital signs and classes

To label data with actual classes representing the patient's situation, global vital sign ranges are used with the advice of a medical professional. The three patient status classes (or labels) are presented as high, normal, and low, along with a specific vital sign. Normal, low and high ranges of vital signs for both diseases are shown in Table 4.

Table 4. Vital sign value range

Vital signs	Normal ranges	Low ranges	High ranges
HR	60 AND 100	Less than 60	Greater than 100
SpO2	94 AND 100	Less than 94	Greater than 100

We combine these ranges with vital signs to determine the output classes. If all relevant vital signs are in low values/ranges, then the corresponding output class will show Low (low means the patient is in a critical condition and requires urgent attention), which corresponds to the patient's situation. If one vital sign is in the low range as mentioned in Table 4 and the other is in the normal range as mentioned in Table 4, then the output class will refer to the low vital sign as the keyword Low (i.e., low heart rate) and ignore the normal vital sign sign. The same applies to the high range (high range means critical patient condition and urgent need, mentioned in Table 4) of vital signs. We also labeled the combination of high and low vital signs with the keywords "Low" and "High" (i.e. "Low HR", "High spO2"). Table 5 provides details related to cardiovascular disease. These classes of patients were determined in consultation with medical experts.

Table 5. Cardiovascular diseases classification model of 7 output classes

Class ID	1	2	3	4	5	6	7
Class Label	Low	Normal	High	Low HR	Low SpO2	High HR	High SpO2

2.4. Regression model

RNN(LSTM), ARIMA, and VAR are used to predict the next 60-second vital signs because the literature review showed that they are powerful methods for modeling and forecasting time series data. RNNs are designed to handle sequential data, making them well-suited for time series forecasting. They can effectively capture patterns and dependencies in the historical data, which can help to improve the accuracy of predictions [15].

ARIMA is a statistical method that models a time series as a combination of past values, trends, and random noise. This method is often used for univariate time series forecasting, such as vital signs, because it can effectively capture trends and seasonality in the data [16]. VAR is a statistical model that considers the relationship between multiple time series variables. It can be used to forecast vital signs by taking into account the relationship between the vital signs and other relevant variables, such as patient demographics, medical history, and other patient data [17].

2.5. Classifiers studies

Different researchers have used different methodologies ranging from linear regression to neural networks for forecasting vital signs [18]. Many authors have compared various data mining methods in the field of medicine for predicting various diseases. Thus, after reviewing the publications of different authors, we found that SVM and Naive Bayes have high predictive accuracy in the medical field for various diseases. Therefore, for our experimental study, we choose the SVM and the Bayes algorithm [19].

The SVM is employed for a classification task. Based on the classification error and the distance between classes, the method modifies the weight vector (w) and bias term (b) throughout the training phase. The weights are changed to move a data point farther away from the margin if its margin is less than 1, indicating that it is within a "margin buffer zone." The learnt weights and bias for each test instance are used to determine the decision score during the classification phase. The method predicts one class if the score is positive; if not, it predicts the other class [20]. SVM seeks to identify a hyperplane that maximizes the margin between the classes while best separating them. This description adequately conveys [21].

Based on instance counts and conditional probabilities for features given each class, it calculates prior probabilities for each class during training [22]. In classification, it multiplies prior probabilities with matching likelihoods for its features for each test occurrence. The prediction is made for the class with the highest product [23]. The algorithm takes into account class-level feature independence. While the main steps are covered in this description, real implementation necessitates resolving problems like zero probabilities and various feature kinds.

2.6. Evaluation

To evaluate the proposed methodology, some standard metrics are applied to assess the performance of each algorithm. These metrics are precision, recall, F-measure, MSE [24]. precision (P): precision measures the proportion of true positive predictions among all positive predictions made by the model.

$$P = TP/(TP + FP) \tag{1}$$

where:
TP=true positives (correctly predicted positive instances)
FP=false positives (incorrectly predicted positive instances)
Recall (R): recall calculates the proportion of true positive predictions among all actual positive instances.

$$R = TP / (TP + FN) \tag{2}$$

where:
TP=true positives (correctly predicted positive instances)
FN=false negatives (actual positive instances missed by the predictions)
F-measure (F1 score): The F-measure combines precision and recall into a single metric, providing a balance between the two.

$$F - score = 2 * (precision * recall) / (precision + recall) \tag{3}$$

MSE: MSE calculates the average of the squared differences between predicted and actual values. The formula for the mean squared error is MSE=Σ(yi−pi)2n, where yi is the ith observed value, pi is the corresponding predicted value for yi, and n is the number of observations. The Σ indicates that a summation is performed over all values of I [25]. These equations provide a quantitative way to assess the performance of each algorithm within the context of the remote patient management system [26].

3. RESULTS AND DISCUSSION

In this section, regression models (i.e., RNN(LSTM), ARIMA, and VAR) are trained on 5,000 samples at 10 millisecond intervals for each vital sign (i.e., heart rate and SpO2) of heart failure. The trained models of each vital sign are then used to predict the next 60 seconds of the same vital sign. These vital sign predictions are used as input to the classification methods (such as SVM, Nave Bayes with output class) already described in section 3.4. Therefore, since this is a supervised learning method, we evaluated the results using data mining methods (i.e., the regression methods and classification methods used in this study) against the results of standard scoring measures (i.e., precision, recall, and f-measure. equal to 1 means that the prediction accuracy is high).

The two-step algorithm for vital sign prediction and classification begins by forecasting the next 60 seconds of vital sign measurements using time-series prediction models such as LSTM, VAR, or ARIMA. Initially, the data is split into training and testing subsets. Using the training data, the selected prediction model is trained and subsequently used to predict the next 60 seconds of measurements based on the test data. Following this, the predicted 60-second data is leveraged to extract pertinent features (mean and variance), which are then fed into a classification model SVM or Naive Bayes. This classifier, trained on the extracted features from the training data, then predicts into one of the seven classes, ranging from low to high vital sign levels. The outlined algorithm provides a structured approach to first forecast short-term vital sign changes and subsequently classify them into predefined categories, serving as a comprehensive method for vital sign monitoring and alerting. Further, we describe in detail the obtained results and compare different prediction methods used in the algorithm.

3.1. Classification using RNN(LSTM)

In this section, the results of RNN(LSTM) and two classification methods are shown in Table 6 and Figure 5. In Table 6, the first column shows two types of classification algorithms (i.e. SVM, Naive Bayes) with heart failure vital signs. The second column shows the results of RNN estimation methods (LSTM) (i.e. MSE). A low MSE value means that the RNN(LSTM) predicts vital signs with a minimum error rate and high accuracy. The RNN(LSTM) values are repeated for each classification method to analyze classification performance for the same predictors (i.e. RNN predictors (LSTM) of vital signs). The third column presents the accepted classification methods (e.g. SVM, Naive Bayes) with precision, recall, and F-score. The third column presents those classification methods that predicted high accuracy. A high value (i.e., up to 1) for precision, recall, and F-score indicates that the classification method predicted heart failure with high accuracy. We focused on the results of F-measures for evaluating data mining methods. The results shown in Table 6 are also presented in chart form. These results are based on an RNN model (LSTM) that is trained on 5,000 samples at 10 ms intervals to predict data for the next 60 seconds. The result shows that SpO2 gives a minimum error rate of up to 1.664, hence SpO2 has a high accuracy. In addition, HR showed the error rate of 1,752 points. The performance score column in Table 6 contains three different types of evaluation methods (i.e. precision, recall, and F-score) for evaluating classification methods. According to the result in Table 6, SVM achieved a high prediction accuracy score of 0.84. Naive Bayes predicts heart failure by f-test with a predictive accuracy of up to 0.81, which is also satisfactory.

Table 6. Classification of 60 seconds forecasted value by RNN(LSTM) of CVD

Classifier		RNN(LSTM)	Performance metrics		
	Vital sign	MSE	Precision	Recall	F measure
Naïve Bayes	HR	1.919	0.76	0.88	0.81
	Spo2	1.878			
SVM	HR	1.752	0.77	0.92	0.84
	Spo2	1.664			

3.2. Classification using ARIMA

By applying the same process to the ARIMA predictive values, we achieved an improved result. The results of ARIMA and all two classification methods are shown in Figure 6 and in Table 7. By carefully monitoring the MSE, the ARIMA model predicts SpO2 with a minimum error estimate of up to 0.114, which is a satisfactory result. According to the result the SVM method achieved a high F-measure prediction accuracy score of up to 0.86.

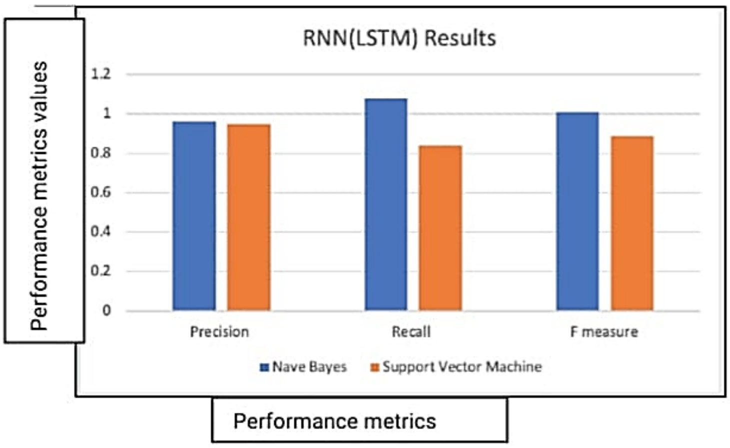

Figure 5. Classification of 60 seconds forecasted value by RNN(LSTM) of CVD

Table 7. Classification of 60 seconds forecasted value by ARIMA of CVD

Classifier		ARIMA	Performance metrics		
	Vital sign	MSE	Precision	Recall	F measure
Naïve Bayes	HR	1.729	0.77	0.86	0.82
	Spo2	1.688			
SVM	HR	1.262	0.88	0.84	0.86
	Spo2	1.114			

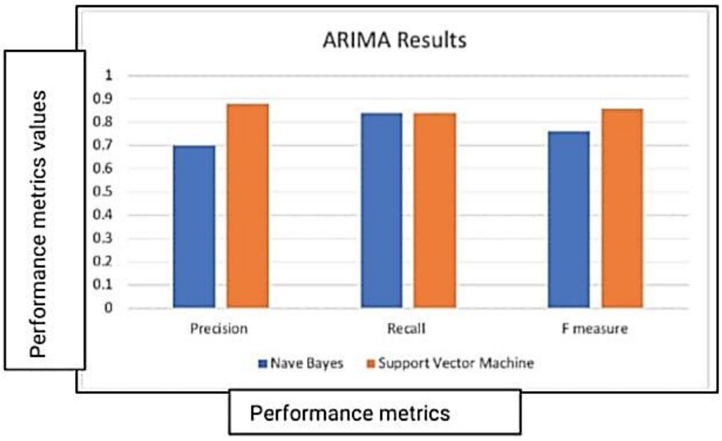

Figure 6. Classification of 60 seconds forecasted value by ARIMA of CVD

3.3. Classification using VAR

We also implemented the VAR model and evaluated its effectiveness. The results of VAR and all two classification methods (i.e. SVM, Naive Bayes) are shown in Figure 7 and in Table 8. If we take a close look at the MSE, the error results do not improve with the VAR model. The minimum MSE score is up to 1.664 and the maximum is up to 1.919. In addition, you may notice that performance scores have not improved.

Table 8. Classification of 60 seconds forecasted value by VAR of CVD

Classifier		ARIMA	Performance metrics		
	Vital sign	MSE	Precision	Recall	F measure
Naïve Bayes	HR	1.529	0.75	0.77	0.75
	Spo2	1.623			
SVM	HR	1.562	0.87	0.77	0.82
	Spo2	1.126			

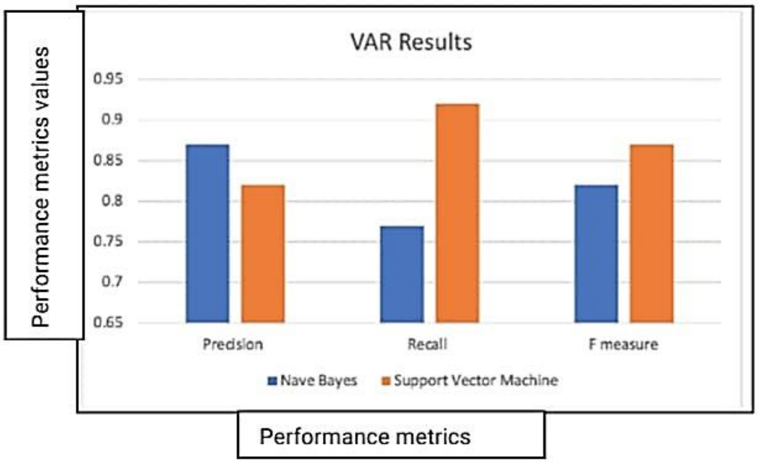

Figure 7. Classification of 60 seconds forecasted value by VAR of CVD

3.4. Comparison analysis

Figure 8 shows a comparison of the above results in this paper. Here, the correlation between ARIMA and SVM algorithms shows a higher F value of 0.86 than the others. Therefore, it is effective to use the ARIMA algorithm to find the relationships of the data points and labels and to use the SVM algorithm to cluster these data.

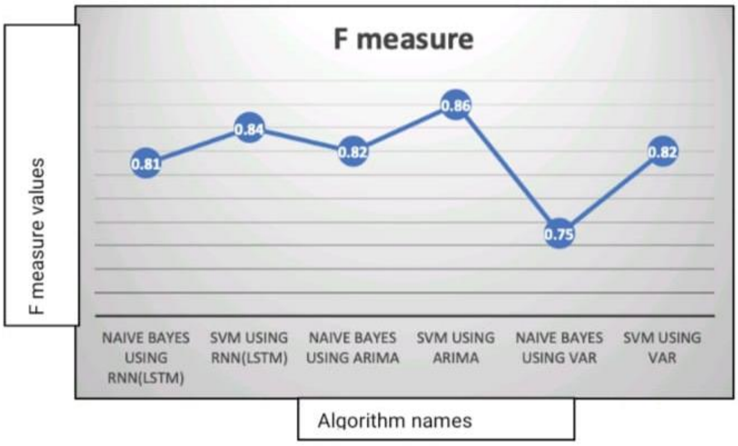

Figure 8. Algorithm correlation results

Table 9 shows the comparison of our research results with the results of [27]. As can be seen in this table, in [27] used the correlation of polynomial degree two, three, four regression algorithms and Naive Bayes and SVM classification algorithms and the best F score was polynomial degree two and SVM algorithm as 0.83 [27]. In our study, a better F-measure of 0.86 was achieved in the correlation of ARIMA and SVM algorithms of this indicator.

Table 9. Comparison results

This paper (model name)	F measure	Reference paper (model name)	F measure
Naïve bayes using RNN(LSTM)	0.81	Naïve bayes using polynomial degree two	0.08
SVM using RNN(LSTM)	0.84	SVM using polynomial degree two	0.83
Naïve bayes using ARIMA	0.82	Naïve bayes using polynomial degree three	0.15
SVM using ARIMA	0.86	SVM using polynomial degree three	0.1
Naïve bayes using VAR	0.75	Naïve bayes using polynomial degree four	0.001
SVM using VAR	0.82	SVM using polynomial degree four	0.001

4. CONCLUSION

The results of this study indicate that a device based on ESP32 and a mobile application using Flutter framework in Dart language were able to effectively collect vital sign data using heart beat and SpO2 sensors. As part of this study, a module was created that predicts what the heart rate and Spo2 vital signs will be in the next 60 seconds. For data mining, the Python programming language with NumPy, scikit-learn, StatsModels, and Matplotlib libraries were used. The following regression machine learning algorithms were used to predict the data: RNN(LSTM), ARIMA, and VAR. The results of these algorithms were used as input to the SVM and Naive Bayes algorithms used to classify vital signs into classes. Thus, the best correlation of the above regression and classification algorithms was found. The F-measure metric used to evaluate the performance of a machine learning model was chosen as the main measure when comparing algorithms.

The results show that the RNN model (LSTM) had a low MSE, indicating high accuracy in predicting vital signs. The ARIMA model was able to achieve an improved result compared to RNN(LSTM) and predicts SpO2 with a minimum error estimate of up to 0.114. The Naïve Bayes method achieved a high f-measure prediction accuracy of up to 0.76 when used with the ARIMA model. The best F index of 0.86 was achieved in the correlation of SVM and ARIMA algorithms. It is important to note that the results of the study are based on a limited set of data and further studies are needed to confirm these results.

ACKNOWLEDGEMENTS

This work was funded by Committee of Science of Republic of Kazakhstan AP09260767 "Development of intellectual information-analytical system for assessing the health of students in Kazakhstan" (2021-2023).

REFERENCES

[1] "Preface of the 22nd conference of open innovations association FRUCT," in *2018 22nd Conference of Open Innovations Association (FRUCT)*, May 2018, pp. VII–X, doi: 10.23919/FRUCT.2018.8468266.
[2] M. W. Woo, J. Lee, and K. Park, "A reliable IoT system for personal healthcare devices," *Future Generation Computer Systems*, vol. 78, pp. 626–640, Jan. 2018, doi: 10.1016/j.future.2017.04.004.
[3] N. M. Turab, J. I. Al-Nabulsi, M. Abu-Alhaija, H. A. Owida, M. Alsharaiah, and A. Abuthawabeh, "Towards fostering the role of 5G networks in the field of digital health," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, p. 6595, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6595-6608.
[4] D. Gupta, A. Parikh, and R. Swarnalatha, "Integrated healthcare monitoring device for obese adults using internet of things (IoT)," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 2, p. 1239, Apr. 2020, doi: 10.11591/ijece.v10i2.pp1239-1247.
[5] B. Al-Duwairi, W. Al-Kahla, M. A. AlRefai, Y. Abedalqader, A. Rawash, and R. Fahmawi, "SIEM-based detection and mitigation of IoT-botnet DDoS attacks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 2, p. 2182, Apr. 2020, doi: 10.11591/ijece.v10i2.pp2182-2191.
[6] S. A. Zende, T. Kulkarni, S. Yadav, A. Biradar, and A. Devare, "Live tracking of saline for betterment of patient," *International Journal of Reconfigurable and Embedded Systems (IJRES)*, vol. 9, no. 3, p. 178, Nov. 2020, doi: 10.11591/ijres.v9.i3.pp178-182.
[7] V. Mach, M. Adamek, J. Sevcik, J. Valouch, and K. Barcova, "Design of an internet of things based real-time monitoring system for retired patients," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1648–1657, Jun. 2021, doi: 10.11591/eei.v10i3.2699.
[8] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, "Smarter traffic prediction using big data, in-memory computing, deep learning and GPUs," *Sensors*, vol. 19, no. 9, p. 2206, May 2019, doi: 10.3390/s19092206.
[9] G. Xu, Y. Shi, X. Sun, and W. Shen, "Internet of things in marine environment monitoring: A review," *Sensors*, vol. 19, no. 7, p. 1711, Apr. 2019, doi: 10.3390/s19071711.
[10] Z. Dubirova and B. Mendybayev, "Managing the transformation of relations between the state, the city and citizens on the example of iKomek109 - Digital service model of nur-sultan city," in *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Apr. 2021, pp. 1–5, doi: 10.1109/SIST50301.2021.9465985.
[11] A. V. L. N. Sujith, G. S. Sajja, V. Mahalakshmi, S. Nuhmani, and B. Prasanalakshmi, "Systematic review of smart health monitoring using deep learning and artificial intelligence," *Neuroscience Informatics*, vol. 2, no. 3, p. 100028, Sep. 2022, doi: 10.1016/j.neuri.2021.100028.
[12] D. Asif, M. Bibi, M. S. Arif, and A. Mukheimer, "Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization," *Algorithms*, vol. 16, no. 6, p. 308, Jun. 2023, doi: 10.3390/a16060308.
[13] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, and K. Kotecha, "Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease," *PeerJ Computer Science*, vol. 7, p. e369, Feb. 2021, doi: 10.7717/peerj-cs.369.
[14] L. P. Malasinghe, N. Ramzan, and K. Dahal, "Remote patient monitoring: a comprehensive study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 57–76, Jan. 2019, doi: 10.1007/s12652-017-0598-x.
[15] Y. Z. Abd El Gawad, M. I. Youssef, and T. M. Nasser, "Medical system based on thermal optical system and neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, p. 3796, Aug. 2023, doi: 10.11591/ijece.v13i4.pp3796-3804.
[16] Y. Zaitin, O. Postolache, and M. Mansurova, "Design of IoT based patient health monitoring system," in *Asian Conference on Intelligent Information and Database Systems*, 2022, pp. 751–762.
[17] P. M. Lo and A. A. Aziz, "A FPGA threshold-based fall detection algorithm for elderly fall monitoring with verilog," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2477–2487, Oct. 2021, doi: 10.11591/eei.v10i5.3152.
[18] J. H. Majeed and Q. Aish, "A remote patient monitoring based on WBAN implementation with internet of thing and cloud server," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1640–1647, Jun. 2021, doi: 10.11591/eei.v10i3.1813.

[19] A. N. Ejin, H. T. Yew, M. Mamat, F. Wong, A. Chekima, and S. K. Chung, "Internet of things based real-time coronavirus 2019 disease patient health monitoring system," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, p. 6806, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6806-6819.
[20] A. Ramachandran, A. Ramesh, A. Sukhlecha, A. Pandey, and A. Karuppiah, "Machine learning algorithms for fall detection using kinematic and heart rate parameters-a comprehensive analysis," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, p. 772, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp772-780.
[21] H. Jamali *et al.*, "Racial disparity in oxygen saturation measurements by pulse oximetry: evidence and implications," *Annals of the American Thoracic Society*, vol. 19, no. 12, pp. 1951–1964, Dec. 2022, doi: 10.1513/AnnalsATS.202203-270CME.
[22] A. Scirè, F. Tropeano, A. Anagnostopoulos, and I. Chatzigiannakis, "Fog-computing-based heartbeat detection and arrhythmia classification using machine learning," *Algorithms*, vol. 12, no. 2, p. 32, Feb. 2019, doi: 10.3390/a12020032.
[23] A. Alamsyah, M. Subito, M. Ikhlayel, and E. Setijadi, "Internet of things–based vital sign monitoring system," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, p. 5891, Dec. 2020, doi: 10.11591/ijece.v10i6.pp5891-5898.
[24] S. Venkatraman, S. Parvin, W. Mansoor, and A. Gawanmeh, "Big data analytics and internet of things for personalised healthcare: opportunities and challenges," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, p. 4306, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4306-4316.
[25] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor completion algorithms in big data analytics," *arXiv*, vol. 13, no. 1, 2017.
[26] D. Collier, "Remote monitoring of blood pressure in patients with high blood pressure during the COVID-19 pandemic," *ISRCTN*, 2021, doi: 10.1186/ISRCTN16393332.
[27] S. L. Javan, M. M. Sepehri, M. L. Javan, and T. Khatibi, "An intelligent warning model for early prediction of cardiac arrest in sepsis patients," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 47–58, Sep. 2019, doi: 10.1016/j.cmpb.2019.06.010.

BIOGRAPHIES OF AUTHORS

Yerlan Zaitin is is currently pursuing my passion for artificial intelligence and big data as a second-year Ph.D. student at the prestigious Al-Farabi Kazakh National University. Born and raised in Kazakhstan, he demonstrated exceptional aptitude for technology and a deep curiosity for the field of AI from an early age. From the onset of my academic journey, he displayed drive for knowledge and a thirst for understanding the intricate workings of the digital world. He completed my undergraduate studies in Computer Science at Al-Farabi Kazakh National University, where he consistently ranked at the top of his class. During this time, he actively engaged in research projects, further fueling his passion for AI and Big Data. Motivated by his relentless pursuit of excellence, he made the decision to continue my education at Al-Farabi Kazakh National University, enrolling in the Ph.D. program in artificial intelligence and big data. He can be contacted at email: yzaitin@gmail.com.

Madina Mansurova is the Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of artificial intelligence and big data of Al-Farabi Kazakh National University. The author of over 100 different publications. Her research interests include natural language processing, artificial intelligence, data mining. She can be contacted at email: Madina.mansurova@gmail.com.

Murat Kunelbayev received a bachelor's degree in radiophysics and electronics from Al-Farabi Kazakh National University, Almaty, Kazakhstan in 1997, as well as a Master's degree in Engineering from Al-Farabi Kazakh National University, Almaty, Kazakhstan in 2003. He is currently a Senior Researcher at the Institute of Information and Computational Technologies CS MES RK. He is also an editor and reviewer of many scientific journals. His current research interests include power converters, renewable energy and energy efficiency. He can be contacted at email: murat7508@yandex.kz.

Gulnur Tyulepberdinova graduated from Al-Farabi Kazakh National University in 2002 with a bachelor degree in applied mathematics. In 2004, she graduated from the Master's degree program of Al-Farabi Kazakh National University. Since June 30, 2011, the academic degree of Candidate of Physical and Mathematical Sciences has been awarded. Associate Professor, specialty 05.13.00–Computer science, computer engineering and Management. She has more than 60 scientific papers, including scientific publications with a non-zero impact factor, indexed in the international information resources of Scopus, as well as in publications recommended by the of the Ministry of Education and Science of the Republic of Kazakhstan. Main research areas: system analysis and mathematical programming problems, artificial intelligence and big data. She can be contacted at email: gulnur.tiulepberdynova@kaznu.edu.kz.

Talshyn Sarsembayeva received a bachelor's degree in information systems from Al-Farabi Kazakh National University (2012, Almaty, Kazakhstan), where her thesis focused on building deterministic parsing algorithms for CF languages. Continuing her academic activities, Talshin received a Master of Science degree in information systems from the University of Turan (2014, Almaty, Kazakhstan). Actively participated in various projects, as a junior researcher, research fellow. Currently he is the deputy head of the department of artificial intelligence and big data at Al-Farabi Kazakh National University and a 3rd year doctoral student in the educational program of artificial intelligence in medicine. Talshyn Sarsembayeva is an active member of the Institute of Electrical and Electronics Engineers (IEEE), a renowned professional organization dedicated to the advancement of technology and innovation. IEEE membership demonstrates Talshin's commitment to remaining connected to the global community of professionals in electrical engineering, computer science, and related disciplines. Her research interests include deep learning, soft computing, machine learning, and intelligent systems. She can be contacted at email: talshyn.sagdatbek@kaznu.edu.kz.

Dr. Adai Shomanov obtained his undergraduate and graduate degree from Al-Farabi Kazakh National University. Currently, he works as an Instructor in Nazarbayev University. His research area includes deep learning, parallel computing and artificial intelligence fields. He can be contacted at email: adai.shomanov@nu.edu.kz.