INFORMATION AND ANALYTICAL SYSTEM FOR ASSESSING THE HEALTH STATUS OF STUDENTS

The purpose of our work was to study the effectiveness of using an intelligent information-analytical system to assess the health status of students on the basis of one of the universities in Kazakhstan, the Kazakh National University named after Al-Farabi. For this purpose, a simulation was performed using individual health data 4456 university students. The study involved 47.4% of female students and 52.6% of males; the predominant age ranged from 18 to 21 years (66.9%), and the distribution of students by years of study was almost uniform. For classification were used such as Vector Machine, K - Nearest Neighbour, Random Forest and Naive Bayes supports. The performance metrics chosen to evaluate the use of various prediction algorithms were Specificity, Sensitivity, Accuracy, and Accessibility. It has been established that when using the classifier Support Vector Machine’s Specificity, Sensitivity, Accuracy, and Accessibility scores are at their highest, reaching 97%. The overall performance of the developed intelligent information and analytical system was evaluated using the Reliability parameter. In comparison with other well-known systems for monitoring the health of patients (AmbIGEM and AAL), the system developed by us showed higher reliability (90-95%). In the future, the developed model can be used to expand health monitoring by including external parameters that can also affect the health of students. In addition, it is planned to introduce Deep Learning to monitor the health of students in other educational institutions in Kazakhstan and the world.

Key words: Internet of Things, Ambient Intelligence, Artificial Intelligence, Health monitoring, algorithm, Simulation.

Г.А. Тюлепбердинова1*, Т.С. Сарсембаева1, С.А. Адилжанова1, С.Н. Исабаева2
1Ал-Фараbi Казakh National University, Kazakhstan, Almaty
2Egyptian University of Islamic Culture «Нур-Мубарак», Kazakhstan, Almaty
*e-mail: Gulnur.Tiuleberdynova@kaznu.edu.kz

Студенттердің денсаулық жағдайыңың басқауы ақпараттық-талдау жұйесі

Біздің жұмысқағызымдық мақсаты Қазақстандағы жоғары оқу ортдарының бірі Ал-Фараби атындағы Қазақ ұлттық университетінің басқа ортада студенттердің денсаулық жағдайының басқауы ақпараттық-талдау жұйесін пайдаланудың төмendezілігін зерттеу болды. Осы мақсатта студенттердің 4456 студентінің және денсаулық дерекетерін пайдаланып модельге жүргізілді.
Зерттеуге студенттердің 47,4% және ерлердің 52,6% қатысты; басыңыз 18-ден 21 жасқа дейін (66,9%), ал студенттердің оқу қымдәрі бойынша бәлініуі біркелкі дерлік болды. Классификация үшін Vector Machine, K - Nearest Neighbor, Random Forest және Naive Bayes қоңұрамдары пайдаланылды. Эртурлі болуға арқылы алгоритмдерін пайдалануға бағалау үшін таңдалған өңімділік қорестің әрекетін Ережелік, Сәйімділік, Дәлдік және қол қосындылық болды. Қоңұрда векторлық машинаның классификаторы пайдалану кезінде ережелік, сәйімділік, дәлдік және қол қосындылық үшін жағының 97%-ға жететін ыңғайылы. Жасалған интеллектуалды ақпараттанық-аналитикалық жүйеңің жалпы өңімділігі Сенімділік параметрі әрқылы бөлінеді.

Пациенттердің денсаулығының бакылауынан бәрі көбірек жүйелерімен (AmbIGEM және AAL) салыстырмайтады. Бұл, оның 90-95% қосындылығы әсер етеді. Бұл өңімділік қорестің кезінде студенттердің денсаулығының бөлінеді.

Туынды сөзлер: интернет вещи, окружающий интеллект, искусственный интеллект, мониторинг здоровь, алгоритм, моделирование.

Г.А. Тюлебердинова1*, Т.С. Сарсембаева1, С.А. Адилжанова1, С.Н. Исабаева2
1Қазақстан Гуманитарийлық Университет имени аль-Фараби, Казахстан, г. Алматы
2Қазақстан Исламский университет "Нур-Мубарак", Казахстан, г. Алматы
*e-mail: Gulnur.Tiuleberdynova@kznu.edu.kz

Информационно-аналитическая система оценки состояния здоровья студентов

Целью нашей работы являлось изучение эффективности использования интеллектуальной информационно-аналитической системы для оценки состояния здоровья студентов на базе одного из университетов Казахстана, Казахского Национального университета имени Аль-Фараби. С этой целью была выполнена симуляция с использованием индивидуальных данных о здоровье 4456 учащихся университета. В исследовании участвовало 47,4% студентов женского пола и 52,6 % мужского пола; преобладающий возраст составил от 18 до 21 года (66,9%), а распределение студентов по годам обучения было практически равномерным.

Для классификации использовались такие алгоритмы, как Support Vector Machine, K-Nearest Neighbour, Random Forest и Naive Bayes. Показателями эффективности, выбранными для оценки использования различных алгоритмов прогнозирования, были Specificity, Sensitivity, Accuracy и Accessibility. Установлено, что при использовании классификатора Support Vector Machine значения показателей Specificity, Sensitivity, Accuracy и Accessibility максимальны и достигают 97%. Общая производительность разработанной интеллектуальной информационно-аналитической системы была оцена с использованием параметра надежности (Reliability). В сравнении с другими известными системами мониторинга здоровья пациентов (AmbIGEM и AAL) разработанная система показала более высокую надежность (90-95%). В будущем разработанную модель можно использовать для расширения мониторинга здоровья за счет включения внешних параметров, которые также могут повлиять на здоровье учащихся. Кроме того, планируется внедрить технологию Deep Learning в мониторинг здоровья учащихся в других учебных заведениях Казахстана.

Ключевые слова: Интернет вещи, окружающий интеллект, искусственный интеллект, мониторинг здоровья, алгоритм, моделирование.

1 Introduction

Innovations in communication devices, data processing technologies and computational intelligence contribute to the improvement of personalized healthcare [1, 2, 3]. In particular, these technologies are used to predict mortality in patients with paralytic ileus [4], to provide care to patients with Parkinson’s disease [5], to expand the access of employees
of organizations to expensive drugs and services [6]. Continuous in-depth research using diagnostic methods, modeling and expanding knowledge about diseases can prevent or eliminate many diseases in time [7, 8]. The use of the Internet of things [9, 10] has made an invaluable contribution to the era of personalized monitoring of human health. Individualized medical recommendations for individuals are given using quantum diagnostic tracking [11], wearable medical sensors [12], non-invasive monitors [13], embedded health sensors [14], audio, visual stimuli, and combinations of these methods [15].

Real-time health alerts are very important for prevention, especially if disease detection can minimize negative body reactions and medical costs at an early stage. Diagnosis and timely assistance in various conditions can radically change the trajectory of a patient’s treatment. The results, analyzes and insights from these individual decisions help identify and assess health risks associated with chronic disease and overall health/well-being at an early age. Recently, there has been a significant improvement in technological innovation, improvement of environmental and health sensors for diagnostics and statistical results using Ambient Intelligence (AmI) models [16, 17]. AmI-based healthcare systems now support physicians by providing effective guidance and information that guarantees a more outstanding quality of life while lowering healthcare costs.

AmI is a computational model that does not include traditional input and output media. Instead, actuators and sensors are built into traditional objects in their living conditions, in harmony with people. To perform these actions, AmI uses artificial intelligence (AI) [18-20]. It captures, interprets and adapts the world to the perceived needs of quality knowledge from embedded sensors. AmI’s healthcare integration is user-friendly, customizable, contextual, predictable, ubiquitous, and simple. AmI is based on AI and is actively progressing in areas such as data collection using sensor nodes, robots and immersive interfaces between people and machines.

Recent sensor network innovations such as the Internet of Things (IoT) are revolutionizing cost-effective home and work healthcare services [21, 22]. AmI medical services use assistive technologies and innovations of two important types: complex sensor networks and networks of body regions in intelligent environments [23, 24].

The healthcare infrastructure has many implementations, providing consumers and healthcare providers with options. Real-time patient health monitoring, through which sensors collect information, allows medical equipment to work more efficiently [25]. The patient’s health information is first transmitted to the network and then sent to the server. In a machine-to-machine environment, IoT nodes can interact with people or other machines [26].

The paper [27] presents the results of a study on the development of an improved medical structure for the interaction between the patient and the doctor. In this work, new therapeutic models were built to identify information priorities that could be labeled as patient emergencies. In the proposed framework, complete diagnostic data has been provided to ensure reliable long-term diagnosis of patients. In light of the different distribution of data priority in the real case, their detailed studies have shown that the approach proposed by the authors will improve the quality of data collection and optimize patient waiting times. Ideally, the proposed system saved 20% of the amount of transmitted data and reduced patient waiting times by 75%.

The authors of [28] proposed the concept of integrating artificial (AI) and ambient
intelligence (AmI) into the development of AI -related information and communication technologies using the rules of the information society, superintelligence and several related disciplines, including multi-agent networks and the Semantic Web. The developed concept should develop such concepts as environmentally supported life, e-health, as well as Ambient Intelligence to support medical diagnostics, e-learning and environmentally sustainable intelligence.

The authors of [29] defined the concepts and presented a plan for increasing the autonomy of intelligent machines, sensors, network applications and smartphone applications based on the Internet of Things in healthcare and education. To ensure equal resources, perspectives for everyone in the world have described the features of low-cost technologies such as Smart Health Tracking, Smart Schooling and Smart Income Networks in the proposed structure. Their work has developed an effective foundation for society using modern hardware, wearable sensors, smartphone apps and web apps.

The work [30] presents a method for monitoring and predicting the state of human health using IoT technology (HSD&P-IoT). This method is designed to eliminate and reduce human intervention in decision making. The method also uses environmental sensors to detect patient behavior, resulting in improved health prediction.

Analyzing the above data, we can conclude that the assessment of health monitoring using advanced digital technologies is at a fairly high level. However, due to the fact that these technologies are modern, they are not well understood in all areas of their possible application. In particular, the field of monitoring and predicting the health of college and university students requires careful study.

The purpose of our work is to study the effectiveness of using an intelligent information-analytical system to assess the health status of students on the basis of one of the universities in Kazakhstan, Al-Farabi Kazakh National University.

As part of the study, we plan to perform the following tasks:

- Determine the architecture of an intelligent information and analytical system for assessing the health status of students (IIAS).

- Set the possibility of using various classifiers (SVM, KNN, RF, NB) in the IIAS system and evaluate their performance by indicators such as Specificity, Sensitivity, Accuracy.

A simulation using data on the health of students from one of the universities in Kazakhstan, Al-Farabi Kazakh National University.

2 Methods and materials

2.1 Intelligent information and analytical system for assessing the health status of students (IIAS)

In our work, we used an intelligent information and analytical system for assessing the health status of students (IIAS), which includes 3 levels: IoT, Cloud and Student health monitoring.

2.1.1 IoT

The IoT layer in IIAS contains several interconnected WSNs (wireless sensor networks) and smart devices to collect the necessary information about the health of students. At this level,
data on the health of students in real time is accumulated using wireless sensors.

Within this level, the first stage is the collection of historical data about the student. At the second stage, data is collected about the current state of health of students (the results of determining the electrocardiogram, electroencephalogram, body temperature, etc.) using installed IoT sensors. In the third stage, each student’s health data is registered in the health monitoring system through certain mobile applications.

2.1.2 Cloud

The information about the health status of students obtained at the first level is stored in the cloud data center. The information includes operational and historical data obtained from medical devices and smart sensors, which are equally important for assessing the state of human health.

The cloud layer in IIAS performs the function of storing/retrieving student health data for further processing. To gain access to information stored in the cloud, a request must be received from the applicant. After that, the authorized user gets access to the cloud data. This user has the right to add or retrieve necessary student health information.

Moreover, two classes of users can create a security system using the cloud robotics database. Class 1 is made up of authorized physicians and observers who have access to shared and independent records, and class 2 is made up of users who need knowledge to produce experimental drugs.

2.1.3 Student health monitoring

This layer processes real-time and historical data stored in the cloud layer.

At the first stage, the necessary data on the health of students is extracted. In the second step, a pre-processing method is applied, which is used to prepare the raw data and adapt the applicability of the data of the corresponding AmI algorithm. Preprocessing is the most important step in the development of an AmI model in order to detect accurate and formatted data. The third step is data normalization, one of the strategies that is performed before classification. This is a hierarchical method that organizes data in tables and eliminates their duplication, as well as various inconsistencies. At the next stage, the process of classifying the health status of students takes place. The alert system is activated in the event of any emergency. Algorithms such as Support are used for classification. Vector Machine (SVM), K - Nearest Neighbor (KNN), Random Forest (RF) and Naive-Bayes (N.B.). It should be noted that the SVM algorithm is the main classification method, and the KNN, RF and NB algorithms are auxiliary.

2.2 SVM classifier

The SVM algorithm aims to determine the best limit of outcomes by moving away from entire classes as much as possible and not depending on other data points. SVM is a classification and data mining algorithm, one of the standard methods that is best for classifying and detecting external results by efficiently forecasting. Since SVM is the main learning algorithm that can be used for classification and regression, both linear and non-linear classification, IIAS performs both linear and non-linear classification.
Provided the data is linearly separable, straight lines can be used to classify two types of student health data. This line is a 2D hyperplane that corresponds to data points -1 and 1. The key problem with using SVM is the intersection of the hyperplane, which is technically a straight line that better defines the two types of sample. The mathematical model for SVM is configured as follows:

\[ h(p) = m*p + c; h(p) = 0 \]  
\[ g(p) = \text{sgm}(h(p)) \]  
\[ F = q*(m^T*p + c) \]

where \( h(y) \) denotes a hyperplane with coordinates \((p, q)\), which is a feature vector; \( m \) and \( c \) are the displacement vectors in equation (1). The value \( g(p) \) defines a decision function in which it \( \text{sgm}(.) \) represents the signum function using equation (2). Since one of the best approaches to automated decision making is a linear classifier, it simply divides the feature space into two non-overlapping regions \( g(-1, 1) \). The feature vector falling into any area belongs to the class \( g \). This vector is classified by the object and marked as acceptable. For obvious purposes, the SVM model singles out and calls the hyperplane the decision function. Using equation (3) the functional margin for SVM classification is obtained, denoted by \( F \).

\[ m = \sum_j (C^j g^j p^j) \]

For all elements of the feature map, the vector \( m \) can be statistically denoted as shown in equation (4). Where \( C^j \) denotes the coefficient of the \( j \)-th feature space. The linear classifier is trained to find the correct position of the separating hyperplane by adjusting its parameters \( m \) and \( c \). This requires a group of classified vectors \( p^j, g^j \) in which the \( j \)-th vector of the set is equal \( p^j \) to and a known class \( g^j \in -1, 1 \). For any learning process, all learning vectors still behave like a linear combination in a linear SVM, as represented by equation (5).

\[ h(p) = \sum_j C^j g^j (p^j, p) + c \]

where \( h(p) \) denotes a linear combination of all training vectors in a linear SVM that can be considered for classifying a non-linear SVM. The options for obtaining displacement vectors and coefficient vectors differ from linear SVM to non-linear SVM. In a non-linear SVM, the need for maximum margin arises from the non-linearity of the training sets. To perform the determination of the hyperplane causing the least disturbance, the quadratic equation (6) is used:

\[ \min(C^j, c, S^j) = \frac{1}{2} ||m||^2 + R(\sum_j S^j)^n \]

where \( (\forall) h'(p) \geq -S^k \), in which the value of the marginal reserve variable \( S^k \) lies between \([0,1)\). Equation (6) leads to a soft margin SVM for \( n > 0 \) and a hard margin SVM for \( n = 0 \) or \( R = \alpha \). Equation (6) marginalizes learning vectors that invalidate the solution \( h'(p) \). \( R \) is a fit parameter that tracks the balance between maximum and minimum violations. SVM is characterized by several distinctive properties:
1. Many coefficients are zero and the training vectors can be ignored with the remaining support vectors.

2. Kernel functions can be replaced by inner products and accommodate significant non-linear extensions in feature vectors.

3. The SVM in IIAS can classify many health-related signals in very high-dimensional feature maps because it does not have a depth defect. Given the growing value of health information to schools and parents, consider using apps to facilitate the storage, testing, verification, and sharing of secure personal health information. Digital records will help integrate all important health data on a single platform and create an effective student profile.

The MATLAB toolbox can be used to train SVM classification with output error correction codes (ECO - error correction output). The ECO code is an appropriate context for solving multiclass classification problems. By breaking down the assignment and combining the effects of certain indirect classification actions with multiple binaries, ECO code avoids the obvious problem for multi-functional classification.

2.3 Simulation

The simulation took place on the example of one of the universities in Kazakhstan, the Kazakh National University named after Al-Farabi. Health data of 4456 students were used. The study sample of students from one of the universities of Kazakhstan, Kazakh National University named after Al-Farabi.

Table 1. Characteristics of students

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Category</th>
<th>Percentage participation, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor</td>
<td>Female</td>
<td>47.4</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>52.6</td>
</tr>
<tr>
<td>Age, years</td>
<td>&lt;18</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>18-21</td>
<td>66.9</td>
</tr>
<tr>
<td></td>
<td>22-24</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td>&gt;24</td>
<td>5.2</td>
</tr>
<tr>
<td>Year of study</td>
<td>one</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>25.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>24.1</td>
</tr>
<tr>
<td></td>
<td>four</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>17.1</td>
</tr>
</tbody>
</table>

Analyzing the data given in Table 1, we can conclude that 47.4% of female students and 52.6% of male students participated in the study; the predominant age ranged from 18 to 21 years (66.9%), and the distribution of students by years of study was almost uniform.

The performance metrics chosen to evaluate the use of various prediction algorithms are specificity (Specificity), sensitivity (Sensitivity) and accuracy (Accuracy), which were calculated using a confusion matrix. The confusion matrix generates true positive (TP - true positive), true negative (TN - true negative), false positive (FP - false positive) and false
negative (FN - false negative) values.

\[
Specificity = \frac{TN}{TN + FP} \times 100
\]  
\[
Sensitivity = \frac{TP}{TP + FN} \times 100
\]  
\[
Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \times 100
\]

3 Results

In this study, ten iterations were performed to evaluate the performance of four different classifiers: SVM, KNN, RF and NB. On fig. Figures 1–3 show the values of the Specificity, Sensitivity and Accuracy indicators for the four used classifiers (SVM, KNN, RF and NB) for 10 iterations.

In particular, the values of the Specificity indicator when using the SVM classifier range from 91 to 96%, while the value of the Specificity classifier when using the KNN classifier ranges from 84 to 90%, when using the NB classifier - from 85 to 90%, and when using the classifier RF - from 80 to 87%.

As for the Sensitivity indicator, when using the SVM classifier, its values range from 92 to 97%; when using the KNN classifier - from 87 to 95%; when using the NB classifier - from 85 to 94%, and when using the RF classifier - from 82 to 91%.

SVM classifier, the Accuracy score varies from 91 to 96%; when using the KNN algorithm - from 85 to 94%; when using the NB indicator - from 87 to 91%, and when using the RF algorithm - from 82 to 90%.

Figure 1: Data Availability Ratio
In addition to evaluating the performance of the classifiers, the overall volume of implementation of the entire architecture of the student health assessment process was analyzed. The 3-layer IIAS architecture was only considered in a real-time implementation if the proposed structure could provide better classification performance and flexibility. Because IIAS works directly with real-time student data, it was critical to test the structure for scalability and latency. All this was done to achieve a high level of student health monitoring and management reliability. In addition, a scalability assessment of the model was performed to determine the ease of expansion of the network, taking into account the growth of the population and customers of the system. On fig. 1 shows the values of the accessibility factor (AR - Accessibility Ratio) for the used classification algorithms. The data availability ratio was calculated by providing a stochastic number of requests addressed to IIAS. Five evaluation iterations were performed on the proposed basis.

Analyzing the data shown in fig. 1 it can be stated that, as in the case of indicators Specificity, Sensitivity and Accuracy, the highest values of the availability coefficient are obtained when using the SVM classifier (92-95%). When classifiers KNN, NB and RF are used, the availability coefficient values are significantly lower and do not exceed 92%.

4 Discussion

The intellectual information-analytical system developed by us for assessing the state of health of students allows real-time monitoring of the state of health of students and, if necessary, promptly making a decision on the intervention of parents or doctors. Let’s compare the system developed by us with similar systems in order to assess its reliability.

The authors of [31] developed the AmbIGEM (Ambient Intelligent Geriatric Management), which uses innovative eco-smart devices with wearable sensors and new Bluetooth low energy technologies. The performance and cost-effectiveness of this method were calculated using a pragmatic step-by-step test in a real hospital setting. The researchers argued that the proposed model had many significant advantages. In particular, this model allows you to track the health status of numerous patients located in several locations. The authors proposed ingenious methods of using pressure pads on beds or pressure sensors on chairs so that the movement of the patient in several places can be automatically tracked. In addition, the authors proposed a new solution to reduce the frequency of false positives.

On fig. 2, the reliability of the IIAS system we developed was compared with existing health monitoring models (AmbIGEM and AAL). The entire dataset was divided into five subsets to evaluate the reliability factor.

Ambient program to assess the health status of patients. assisted Living (AAL). The AAL program allows you to manage behavior and monitor the patient in real time. The AAL program uses an advanced wireless body sensor network to promote independent living and personalized healthcare. The modern approach to software defined network (SDN) traffic management presented in [34] has shown the flexibility and efficiency of management. The SDN Controller intelligently processes and routes the multitude of complex data collected for the body and climate area network. SDN monitors traffic flow and exchanges traffic flow rules with cameras, wearables, and other mobility and routing control technologies.

Analyzing the data shown in fig. 2, we can conclude that the IIAS system developed by us shows higher values. than the AmbIGEM and AAL systems. In particular, reliability values
Information and analytical system ...

Figure 2: Safety factor

for IIAS range from 90 to 95%, while reliability values for AmbIGEM and AAL do not exceed 91%.

5 Conclusion

This study proposed a new health monitoring system university students with intelligent environment and IoT features to improve student achievement. The proposed structure of the intelligent information and analytical system for assessing the health status of students (IIAS) had three levels: the IoT level, the cloud level, and the level of student health monitoring. The study implemented four different supervised learning algorithms (SVM, KNN, NB and RF) to integrate Ambient Intelligence into student health monitoring. It was found that the SVM classifier showed the best performance (96-97%) using the indicators Specificity, Sensitivity and Accuracy. The network was evaluated by the data availability factor and compared with various algorithms. As with the Specificity, Sensitivity, and Accuracy metrics, the highest Accessibility Ratio ) are obtained using the SVM classifier (92-95%). When classifiers KNN, NB and RF are used, the availability coefficient values are significantly lower and do not exceed 92%.

The overall performance of IIAS was evaluated using the Reliability parameter and showed the highest value (90-95%) compared to existing (AmbIGEM and AAL) models (less than 91%). In the future, the model we have developed can be used to expand health monitoring to include external parameters that may affect student health. In addition, it is planned to introduce Deep Learning in student health monitoring.
Acknowledgment

This research was funded by grant of Ministry of Education and Science of the Republic of Kazakhstan №AP09260767 “Development of an intelligent information and analytical system for assessing the health status of students in Kazakhstan” (2021-2023).

References


