Advances in Digital Technologies
J. Mizera-Pietraszko et al. (Eds.)
IOS Press, 2017

doi:10.3233/978-1-61499-773-3-109

Mobile-Bayesian Diagnostic System for Childhood Infectious Diseases

Precious IHEME[a,1] Nicholas OMOREGBE[a] Sanjay MISRA[a] Davies ADELOYE[a] and Adewole ADEWUMI[a]
[a] Center of ICT/ICE Research, CUCRID Building, *Covenant University, Ota, Nigeria*

Abstract. About 5.9 million children under the age of 5 died in 2015, Preterm birth, delivery complications and infections source a great number of neonatal deaths. the Sustainable Development goals (SDGs) 3.2 is to end preventable deaths of newborns and children under 5 years of age, with a target to reduce neonatal mortality to at least 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births in all countries. However quality and accessible healthcare service is essential to achieve this goal whereas most undeveloped and developing countries still have poor access to quality healthcare. with the emergences on mobile computing and telemedicine, this work provide diagnostics alternative for childhood infectious diseases using Naïve Bayesian classier which has been proven to be efficient in handling uncertainty as regards learning of incomplete data. In this research, sample data was collected from hospitals to model a pediatric system using Naïve Bayes classifier, which produce a 70% accuracy level suitable for a decision support system. The model was also integrated into a SMS platform to enable ease of usage.

Keywords. Childhood diseases, Diagnostic system, Machine learning, Mobile computing, Naïve Bayesian, SMS-based system

1. Introduction

Mobile computing which is a technology that allows the use of a central information wirelessly on portable devices [1]. This technology has enabled the used of amenities which were either unavailable or only accessible to a few. For instances, in disseminating health information and services, the mobile platform has proven its advantage [2-5]. Mobile phones are at this time being used effectively in Africa and in other developing countries to disperse information in the areas of health, agriculture, and as a tool for commerce especially through text messaging.

The rate of child mortality especially the undeveloped and developing countries emphasis the need for health solutions that is readily available and affordable. Organizations such as the SDGs and UNICEF have currently carrying out project to solve this problem [6]. Most childhood diseases, which lead to mortality, are preventable and treatable. Diseases such as malaria, pneumonia diarrhea and measles have a significant consequence on the rate of child mortality. Studies show that sub-Saharan Africa will have 33% of the births and 60% of the deaths in 2030, compared with 25% and 50% in 2013, respectively [7]. Children from poor families are more

[1] Corresponding Author, Sanjay Misra, Department of Electrical and Information Engineering, Covenant University, Ota, Ogun State, Nigeria; E-mail: sanjay.misra@covenantuniversity.edu.ng.

exposed to these diseases due inadequate social amenities in their environment [8]. Also the ratio of pediatrician to patient is relative low, which could lead to delayed diagnosis and consequently death of the child [9].

Machine learning algorithms study pattern within a set of data, these patterns are then used in building models for prediction and decision making purposes [10]. Naïve Bayesian algorithm is a classification algorithm under machine learning, which learns via the assumption of independence between the variables within a given class [11]. Despite this assumption, Naïve Bayes has proven its efficiency in solving complex problem [12] [13]. The missing data and error handling efficiency of Naïve Bayesian classifier have made it suitable and most likely to be used in a domain with uncertainty such as the medical domain.

1.1. Statement of the problem and Objective of the study

Several organizations such as the SDGs and UNICEF are working to reduce the rate of child mortality especially the undeveloped and developing countries emphasis [1], however they are yet to cover the target areas especially those with high mortality rates. Hence the research aims to proffer a diagnostic solution to some common childhood diseases using Naïve Bayes classification model, which runs on a mobile platform.

This work was done through identifying and gathering diagnostic criteria and symptoms of specific childhood diseases, preprocessing the collected data, building the Naïve Bayesian classification model and designed the SMS platform for the system. The preliminary work of this was presented in a conference [14] in which we used WEKA tools to test the dataset collected. In the present work, we are developing our own algorithm and developing a system for the mobile application.

2. Related Work

The evolution of mobile computing has been beneficial in several field of endeavor; several researchers have employed this technology to proffer solution in areas with communication gaps. Castellano et al., in [15] designed a real-time emergency telemedicine system for remote medical diagnosis to demonstrate that it is possible to perform hematological tests in an ambulance in terms of an international normalized ratio (INR) using wireless transmission using hybrid network that enables secure long-distance communication from an ambulance. Results of the tests in the ambulance are such that there were no significant differences between the values obtained from the samples analyzed during travel in the ambulance and those analyzed in the laboratory.

Yoo et al., in [16] designed a personalized mobile patient guide system for a patient-centered smart hospital to address the difficulties that outpatients face in finding hospital facilities, recognizing their daily treatment schedule, and accessing personalized medical and administrative information. Using a real-time location-based outpatient guide system that consists of Bluetooth access points (APs) for indoor navigation, an Android-based guide application, a guide server, and interfaces with the HIS. The average success rate was 67.4%.

Supporting Antenatal Care in developing countries through mobile diagnostic system [17]. This system was designed as a maternal diagnostic system for rural pregnant women to improve antenatal care. The decision support system has short answering time and can classify different diseases for pregnant women with high accuracy.

Several researches have been done to support the use of machine learning in modelling health systems which suitable accuracy levels. Flávio H.D. Araújo et al., Used machine learning to support healthcare professionals in making preauthorization decisions due to the need for round the clock preauthorization reviewer which increases operating expenses [18]. Learning was done using dental set from existing database of a non-profit HIP (health insurance provider). The procedures included pre-processing the data, applying different classifiers and picking the one with the highest accuracy level. From preprocessing, out of 165 attributes, 15 were selected and 12 new ones were added. A decision support mechanism with an accuracy level above 96%.

Kose et al., designed An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance [19] using pair wise comparison method of analytical hierarchical processing (AHP) for weighting actor and attributes, expectation maximization for clustering similar actors, two stage data ware housing for proactive risk calculations visualization tool for effective analyzing and z-score and standardization in order to calculate the risks.

A decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches [20]. The research work came up with a model that provide the critical factors and knowledge associated with cognitive heart failure. The result of the experiment shows a significant difference between the cognitive heart failure patients and others based on their age and gender.

Lai et al., used autoregressive process and support vector machine classifier for Detection of tripping gait patterns in the elderly, which could lead to mortality due to internal bleeding [21]. These methods gave a detection accuracy of 69.57%.

Omar Boursalie et al designed M4CVD [22], a mobile machine learning model for monitoring cardiovascular disease device that sends health data from patients, process the raw data and sends it in a meaningful way to help the care giver make accurate decisions. Svm compare data from the clinical databases and the sensor to classify a patient as either "continued risk" or "no longer at risk". The system successfully classified patients with a 90.5% accuracy level.

e-Doctor is a Web Based Support Vector Machine for Automatic Medical Diagnosis [23]. This system request that the user provides some statistical medical information about a disease, the result will be a prediction about the patients' health condition. The SVM engine was implemented on Matlab, an adapter was also design as an interconnector for the java application and support vector machine (engine).

Umit D. Ulusar in [24] used Naïve Bayesian algorithm for pattern classification, and Minimum Statistics and spectral subtraction for noise attenuation for the detection of recovery of gastrointestinal track motility by observing BSs in auscultation recordings. This approach is suitable for real-time long-term continuous monitoring in clinical environments.

Shahi et al, designed a framework that improves the communication responses and the ability of automated decision support among the heterogeneous systems using Naïve Bayes [25] with a rule-based repository for the management of intelligent buildings. This method proves to solve interoperation problems for intelligent systems.

Bill Karakostas used Naïve Bayes models to predict event generated by objects in an Internet of things environment [26]. The model assigns the highest probability to the correct delay event category in 53.8 % of the cases, which is better than random.

Chang et al., designed Mobile cloud-based depression diagnosis using an ontology and a Bayesian network [27], for obtaining an accurate diagnosis of depression is an important issue in healthcare, prototyped on a mobile agent platform as a proof-of-concept in the mobile cloud.

The study in [28] proposes a Bayesian network model for diagnosis of psychiatric diseases. The system is based on physician inability to identify the precise reason of the psychiatric diseases. The use of Bayesian network led to the identification of the most significant factors that affect most important psychiatric diseases and their correlation.

In [29] a system to determine patient eligibility identification combined with an asthma management system to decrease the time of decision making was designed. The system contains two components that is the asthma detector component, which was designed using Bayesian network and the management component. Out of 13896 patients selected for the study, the Bayesian network identified 1339 as asthma exacerbation positive while 788 has asthmas as their final diagnosis, producing a positive predictive value of 69.9%.

Renata Saraiva et al, designed a medical diagnosis decision support model for gastrointestinal cancer [30], which can be used by general practitioners whenever there is a suspicion that a patient has that type of cancer using Case-Based Reasoning (CBR) and Rule-Based Reasoning (RBR). The model was validated using K-fold cross validation and the paired t-test and the accuracy of the diagnosis increased by 22.92% when compared to a CBR approach not using RBR in case retrieval.

3. Research Methodology

To achieve the objective of this work, data was collected from three hospitals. The data were preprocessed to improve the quality of the data. The data-mining algorithm used is the Naïve Bayes classifier. The model learns a set of preprocessed data collected and the output is classified based on the different diseases of focus for inference purpose.

3.1. Data Acquisition

The diagnostic criteria used in the implementation of this system were obtained from World Health Organization (WHO) fact sheet and the dataset were gathered from the Federal Medical Center, State Specialist Hospital and Hillcrest Hospital all in Lokoja, Kogi State. An interview was also conducted with one of the domain experts. According to WHO fact sheet on childhood health and mortality, leading causes of death in under-5 children are preterm birth complications, pneumonia, birth asphyxia, diarrhea and malaria. About 45% of all child deaths are linked to malnutrition [31]. Six conditions account for about 70% of all child deaths: acute lower respiratory infections, mostly pneumonia (19%), diarrhea (18%), malaria (8%), measles, (4%), HIV/AIDS (3%), and neonatal conditions, mainly pre-term birth, birth asphyxia, and infections (37%) [32] however in the research the first four diseases.

3.2. Naïve Bayesian classification model

Naïve Bayes classification is an algorithm extracted from the Bayes rule. This algorithm assumes that the effect of an attribute value on a given class is independent of the values of the attributes. This is called the class-conditional independence. With appropriate data preprocessing, Naïve Bayes is very effective in programmed diagnosis especially in the medical domain.

Bayes theorem: $P(H/X)=(P(X/H)P(H))/((P(X))$ (1)

Steps involved in Naïve Bayes classification

Step 1: D is a training set and its class variables. Each tuple X is represented by an n-dimensional attribute vector ($X= x_1, x_2\dots x_n$) with n measurement on then attribute ($a_1, a_2\dots a_n$).
Step 2: Given m classes $C_1, C_2,\dots, C_m$ and a tuple X, the classifier will predict that X belong to the class with the highest posterior probability, conditioned on X.

$$P(C_i/X) > P(C_i)P(X) \text{ for } 1 \leq j \leq m, j \neq 1 \quad (2)$$

Step 3: Since P(X) is constant for all the class, only $P(C_i/X)\ P(C_i)$ needs to be maximized.
$P(C_i) = C_iD/D$.
Step 4: To reduce computation in evaluating $P(X/C_i)$, the Naïve Bayes assumption of class-conditional independence is made.

"P(X/Ci) = ∏_(k=1)^nP(Xk/Ci) " (P(X1/Ci), P(X2/Ci),…, P(Xn/Ci)) (3)

Step 5: To predict the class label of X, P(X/Ci)P(Ci) is evaluated for each class Ci.

$$P(X/Ci)\ P\ (Ci)> P(X/Cj)P(Cj) \text{ for } 1 \leq j \leq m, j \neq i. \quad (4)$$

The evaluation criteria for the classifier include the following:
The accuracy is the percentage of correctly classified disease over the total dataset given as

$$(TP+TN)/\ (P+N) \quad (5)$$

The error rate is also known as the misclassification rate given as

$$(FP+FN)/\ (P+N) \quad (6)$$

The sensitivity is the percentage of correct classification for each of the diseases given as

$$TP/P \quad (7)$$

The specificity is the degree of exactness of the classification given as

$$TN/N \quad (8)$$

3.3. Naïve Bayesian Diagnosis system

The Naïve Bayes diagnostic system (as depicted in Fig. 1) designed for this research has the following components.

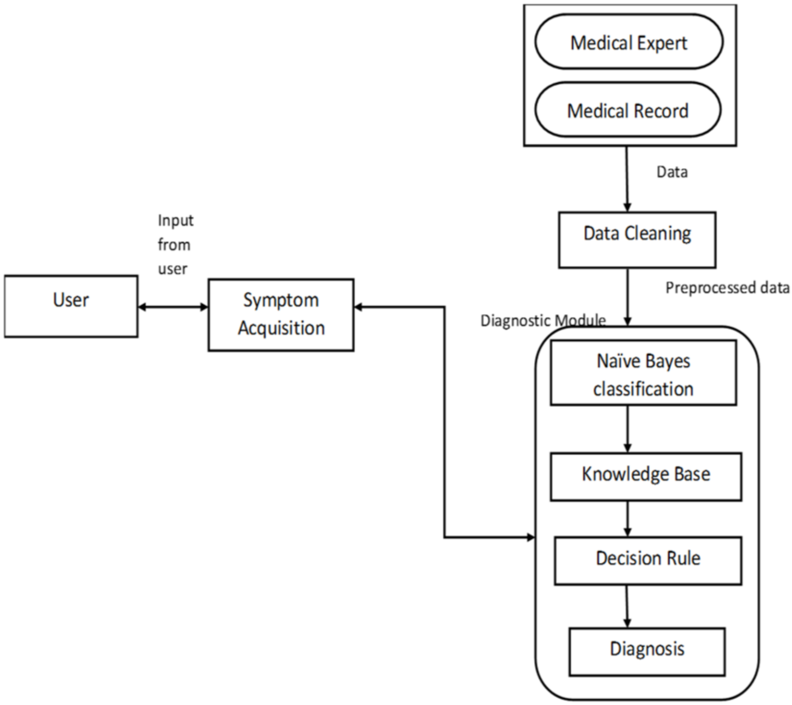

Figure 1. The Naïve Bayes Diagnostic system.

Data module, which contains the raw data, obtains from medical centers and medical experts. The data consist of symptoms presented by patients and the diagnosis from the doctor. The raw data then passes through preprocessing to reduce noise and missing data, and also to convert it to the Naïve Bayes format.

The diagnostic module, which is made up of the Naïve Bayes classification algorithm uses the preprocessed data to build a classifier (model) by learning the dataset. The resulting probabilities from the training are stored in the knowledge base inference purpose. The user symptoms are collected by the symptom acquisition module and sent to the diagnostic module. Using the decision rules and probabilities in the knowledge base, the maximum posterior probability (a suitable diagnosis) is obtained and is sent back to the user.

The SMS- based interface is presented in Fig. 2 and Fig. 3.

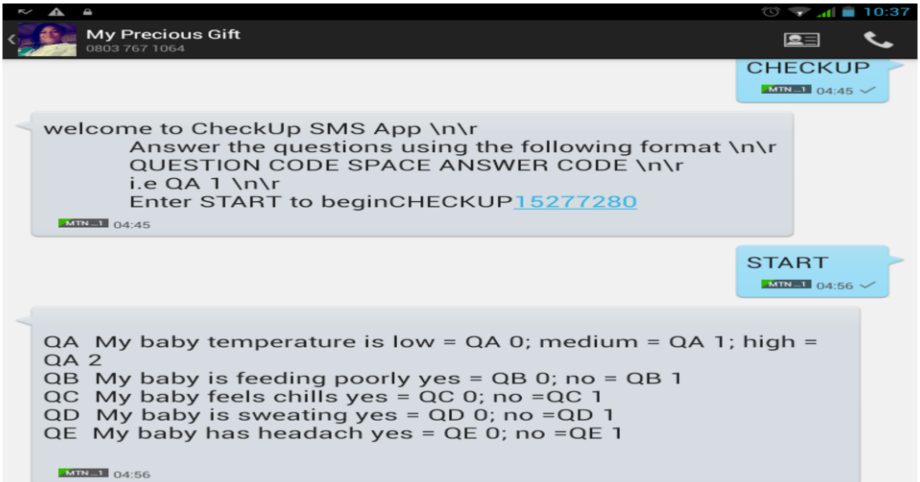

Figure 2. How the user prompts the system.

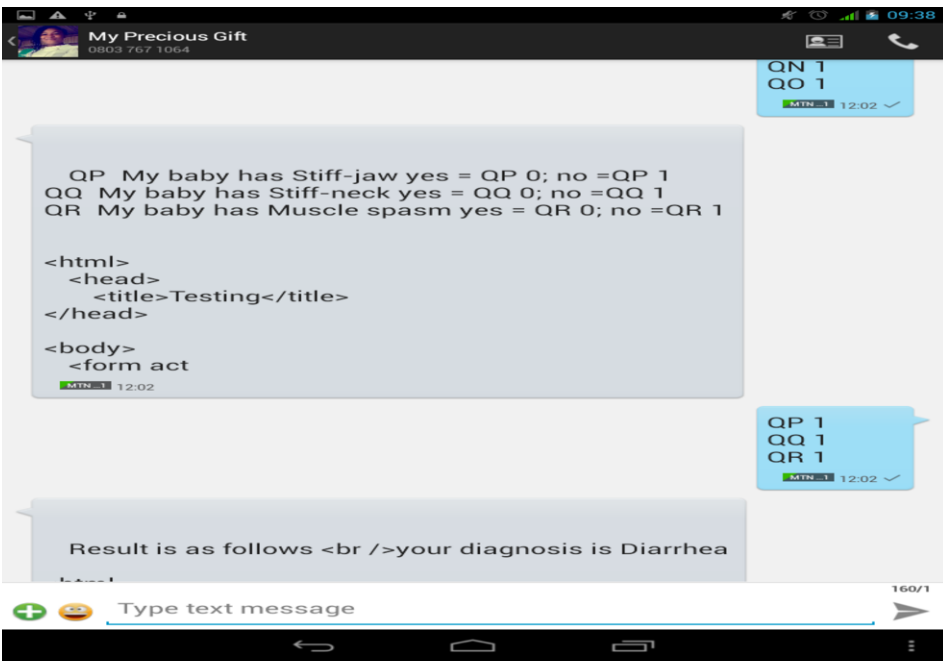

Figure 3. Result of Diagnosis.

The SMS platform makes use of an SMS gateway, which enables communication between the text messages form the user to the Naïve Bayes classification model. A user can prompt the system by sending a keyword to the phone number, which the system receives via the gateway and sends back to the user the instruction for using the system. The user acknowledges this instruction by sending “start” to the system (as depicted in Fig. 2).

A set of question will be sent as soon as the user acknowledges the instructions (as depicted in Fig. 2). The user is required to answer every question from the system. For most of the questions, ‘0’ means yes and ‘1’ means no except in the case of temperature where the number of levels is three that is low, medium and high. If the user fails to answer any of the provided questions, the system will resend that question until the user has satisfied that rule. When the user sends the answers to the last sets of questions, the gateway then sends the answers to the classifier.

When the classifier receives the set of answers from the patient, it extracts the ones that the user answered ‘yes’ to. It then retrieves the conditional probabilities of those symptoms and the prior probabilities of the classes (diseases), these probabilities will then be used to calculation the posterior probability for each of the diseases using the Naïve Bayes formula. The class with the maximum posterior probability will be classified as the diagnosis of the patient. The result will then be sent to the user through the SMS gateway (as depicted in Fig. 3).

4. Results and Discussion

Data (n=1107) use for this research was obtained from the federal medical Centre and the Kogi state specialist hospital in Kogi state, Nigeria. After preprocessing the data was divided into the training dataset (n=580) and the testing dataset (n=527). Table depicts the evaluation of the data using the Naïve Bayes model.

Table 1. Evaluation of the Naïve Bayes model

	Predicted							
	Malaria	Diarrhea	Pneumonia	Measles	Sensitivity	Specificity	Precision	Accuracy
Malaria	215	43	22	0	0.77	0.70	0.74	0.70
Diarrhea	17	107	4	0	0.84	0.87	0.67	
Pneumonia	54	9	48	0	0.43	0.93	0.62	
Measles	4	1	3	0	0	1	0	

The diseases consider for this research were malaria, pneumonia, diarrhea and measles. This evaluation was carried out to determine the accuracy, sensitivity, specificity and precision of the system.

For the evaluation process, the result from the classifier was compared to that of the doctor. The table the number of data that were correctly classified and incorrectly classified based on the doctor’s diagnosis. The Table shows that malaria was correctly classified 215 times out of 280, diarrhea was classified correctly 107 times out of 128, pneumonia was correctly classified 48 times out of 111 and measles was incorrectly classified. The accuracy obtained from the classification model was 70% (as shown in Table 1).

5. Conclusion

The crown of this research was building a diagnostic system for the diagnosis of common childhood diseases. This was achieved by requirements gathering and knowledge acquisition from doctors, health regulatory bodies and relevant patient records from hospitals; detailed literature review to identify possible gaps, methods and tools to solve the problem; building and evaluation of a diagnostic system using Naïve Bayes classification algorithm.

The system developed is to be used by the parent of children between 0-5or medical personnel to obtain an initial diagnosis. Naïve Bayes classification was used to implement the system, which has proven its efficiency in several systems. The system is to reduce the menace caused by delayed diagnosis which leads to child mortality especially in the rural areas as were as contribute the SDGs goal of eliminating child mortality. The classification model was evaluation and its accuracy of 70% shows that it is suitable for initial diagnosis of diseases.

For further research, the existing system will be updated with other diseases using the same method. The system will also be designed with other classification tools, which will be compared to the accuracy of the existing tool.

Acknowledgement

We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation and Discovery (CUCRID).

References

[1] P. Bolliger, *Redpin-adaptive, zero-configuration indoor localization through user collaboration*, Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments. ACM, (2008), 55-60.
[2] C.C. Freifeld, R. Chunara, S.R. Mekaru, E.H. Chan, T. Kass-Hout, A.A. Iacucci, & J.S. Brownstein, Participatory epidemiology: use of mobile phones for community-based health reporting, *PLoS Med*, **7** (2010), 1-5.
[3] J.H.G. Sa, M.S. Rebelo, A. Brentani et al. Georeferenced and secure mobile health system for large scale data collection in primary care, *International Journal of Medical Informatics*, **94** (2016), 91-99.
[4] N. Sun, P.L.P Rau, Y. Li, T. Owen, & H. Thimbleby, Design and evaluation of a mobile phone-based health intervention for patients with hypertensive condition, *Computers in Human Behavior*, **63** (2016), 98-105.
[5] Y. Jiang, S.M. Sereika, D. DeVito, A. Dabbs, S.M. Handler and E.A. Schlenk, Using mobile health technology to deliver decision support for self-monitoring after lung transplantation, International *Journal of Medical Informatics*. **94** (2016) 164-171.
[6] UN-DESA. Transforming our world: the 2030 Agenda for Sustainable Development, Sustainable Development and Knowledge Platform. 2015.
[7] L. Liu, S. Oza, D. Hogan, J. Perin, I. Rudan, J.E. Lawn, S. Cousens, C. Mathers, R.E. Black, Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis, *The Lancet*, **385** (2015), 430-440.
[8] A. Wagstaff, F. Bustreo, J. Bryce and M. Claeson, Child health: reaching the poor, *American journal of public health*, **94** (2004), 726-736.

[9] E.N. Ekure, M.R. Balogun, J.G. Woo, M. Mukhtar-Yola, O.O. Ojo, I.J. Emodi, A.I. Omoigberale, C. Ezechukwu, A.O. Olowu, W.N. Ogala, D.O. Esangbedo, Paediatrician workforce in Nigeria and impact on child health, *Nigerian Journal of Paediatrics*, **40** (2013), 112-118.
[10] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, From data mining to knowledge discovery in databases, *AI magazine*, **17** (1996) 37.
[11] N. Friedman, D. Geiger and M. Goldszmidt, Bayesian network classifiers, *Machine learning*, **29** (1997), 131-163.
[12] Y. Song, A. Kołcz and C.L. Giles, Better Naïve Bayes classification for high ‐ precision spam detection, *Software: Practice and Experience*, **39** (2009) 1003-1024.
[13] P. Tsangaratos and I. Ilia, Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size, *CATENA*, **145** (2016), 164-179.
[14] P.Iheme,N.Omoregbe,S.Misra,F.Ayeni, and D.Adeloye, A Decision Support system for pediatric diagnosis, In Proc of EAI International Conference on Innovations and Interdisciplinary Solutions for Underserved Areas, Held at Dakar, Senegal April 2017
[15] N.N. Castellano, J.A. Gazquez, R.M.G. Salvador, A. Gracia-Escudero, M. Fernandez-Ros and F. Manzano-Agugliaro, Design of a real-time emergency telemedicine system for remote medical diagnosis, *Biosystems Engineering*, **138** (2015), 23-32.
[16] S. Yoo, S.Y. Jung, S. Kim, E. Kim, K.H. Lee, E. Chung, and H. Hwang, A personalized mobile patient guide system for a patient-centered smart hospital: Lessons learned from a usability test and satisfaction survey in a tertiary university hospital, *International journal of medical informatics*, **91** (2016), 20-30.
[17] A. Al Mahmud and D.V. Keyson, *Supporting antenatal care in developing countries through mobile diagnostic system*, Proceedings of the 3rd ACM Symposium on Computing for Development, ACM, (2013), 26.
[18] F.H. Araújo, A.M. Santana and P. Neto, Using machine learning to support healthcare professionals in making preauthorisation decisions, *International Journal of Medical Informatics*, **94** (2016), 1-7.
[19] I. Kose, M. Gokturk and K. Kilic, An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance, *Applied Soft Computing*, **36** (2015), 283-299.
[20] C.S. Son, Y.N. Kim, H.S. Kim, H.S. Park and M.S. Kim, Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches, *Journal of Biomedical Informatics*, **45** (2012), 999-1008.
[21] D.T.H. Lai, R.K. Begg, S. Taylor and M. Palaniswami, Detection of tripping gait patterns in the elderly using autoregressive features and support vector machines, *Journal of Biomechanics*, **41** (2008) 1762-1772.
[22] O. Boursalie, R. Samavi and T.E. Doyle, M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease, *Procedia Computer Science*, **63** (2015) 384-391.
[23] A. Kampouraki, D. Vassis, P. Belsis and C. Skourlas, e-Doctor: A Web based Support Vector Machine for Automatic Medical Diagnosis, *Procedia-Social and Behavioral Sciences*, **73** (2013) 467-474.
[24] U.D. Ulusar, Recovery of gastrointestinal tract motility detection using Naïve Bayesian and minimum statistics, *Computers in biology and medicine*, **51** (2014) 223-228.
[25] A. Shahi, M.N. Sulaiman, N. Mustapha and T. Perumal, Naïve Bayesian decision model for the interoperability of heterogeneous systems in an intelligent building environment, *Automation in Construction*, **54** (2015) 83-92.
[26] B. Karakostas, Event Prediction in an IoT Environment Using Naïve Bayesian Models, *Procedia Computer Science*, **83** (2016), 11-17.
[27] Y.S. Chang, C.T. Fan, W.T. Lo, W.C. Hung and S.M. Yuan, Mobile cloud-based depression diagnosis using an ontology and a Bayesian network, *Future Generation Computer Systems*, **43** (2015) 87-98.
[28] D.I. Curiac, G. Vasile, O. Banias, C. Volosencu, & A. Albu, *Bayesian network model for diagnosis of psychiatric diseases*. In Proceedings of the ITI 31st International Conference on Information Technology Interfaces, (2009) 61-66.
[29] J.W. Dexheimer, T.J. Abramo, D.H. Arnold, K. Johnson, Y. Shyr, F. Ye, F., K.H. Fan, N. Patel & D. Aronsky, Implementation and evaluation of an integrated computerized asthma management system in a pediatric emergency department: a randomized clinical trial, *International journal of medical informatics*, **83** (2014), 805-813.
[30] R. Saraiva, M. Perkusich, L. Silva, H. Almeida, C. Siebra and A. Perkusich, Early diagnosis of gastrointestinal cancer by using case-based and rule-based reasoning. *Expert Systems with Applications*, **61** (2016), 192-202.
[31] World Health Organization. *Children: Reducing Mortality*. World Health Organization. Geneva, 2016.
[32] World Health Organization. *Causes of child mortality*. World Health Organization. 2013.