Case-Based Reasoning Diagnostic System for Antenatal Research Database

https://doi.org/10.3991/ijoe.v18i07.30353

Muhammad Kamran(✉), Ahmed Abdul-Aziz Almaghthawi
College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia
mkkamran@uj.edu.sa

Abstract—E-health is an exciting area of research with immense potential for benefit of mankind. The research in e-health domain is of such a nature that it provides an opportunity for researchers from different fields to work together for collective research benefit. Antenatal care is a critical issue for the health of the mother and new born. There are multiple complications which can arise during the pregnancy. The availability of all the relevant data for effective diagnosis and treatment is very important for every case of antenatal care. In this paper, we have proposed an efficient Intelligent Electronic Medical Record system for antenatal care and used Case-Based Reasoning to predict diagnosis of the patient in noisy environment. The experimental results obtained have confirmed the soundness of our approach. These results have also shown that performance of e-health can be enhanced when system development is a collective effort of various stakeholders including medical professionals, software developers and researchers.

Keywords—case-based reasoning, e-health, clinical decision support system

1 Introduction

INFORMATION technology (IT) is quite effectively used to provide e-health services all over the world. The professionals form health and IT – when working together – are making a large impact both in research as well as healthcare areas. Electronic Medical Records (EMR) systems are quite widely adopted to record patient history. US President Obama promised, “We will make the immediate investments necessary to ensure that within five years all of America’s medical records are computerized.” The incentive amount for adopting EMR systems at a hospital is $11 million. The ultimate goal is “the utilization of a certified electronic health record for each person in the United States by 2014.” In this context, some serious research efforts have been made to facilitate the development of techniques and products for clinical decision support systems (CDSS) [1], medical data mining [2], [3], privacy preserving [4], ownership protection [5], [6] and transfer of medical data over computer networks [7].

Antenatal care is one of the most important health areas and information and communication (ICT) has been used to help medical practitioners and scientific researchers to work for better care of the mother and child [8]-[10]. Antenatal care needs some special care because there are a lot of complications associated during pregnancy for the patient. The two main reasons; the medical practitioners have a very busy schedule, and the missing information about the patient makes it difficult to diagnose the patient for all the complications simultaneously. So a model has to be developed that facilitates the doctors and physicians to diagnose a patient while keeping in view all the possible complications at a time.

An effective IT based solution for antenatal care can play a vital role in making the lives of the people better. The doctors can better diagnose and treat the patient and scientific researchers can better employee their skills to make effective e-health projects. The major problems in such developing such a system are that the complete information of the patient is not available and also the available information is heterogeneous in nature. These problems are severe in developing countries. Another problem is that in developing countries the patient information is not stored electronically due to lack of technology development and busy schedule of the doctors. The major bottleneck is to transfer the manual records into an electronic form.

The Electronic medical records (EMR) are built specially for transferring the data from the manual system into a database from it can be retrieved time to time when required. The EMR developed specifically for antenatal care is different in the sense that it has to store all the information from patient ID to lab investigation. The information collected is very sparse in nature because all the patients not necessarily are examined for everything. So EMR needs to handle all such information in such a way that each and everything is noted without putting too much burden on the doctor. Once such an EMR has been developed then there is a need of such a system which helps the doctor to diagnose the patient accurately and efficiently. The system should be able to; work when information is not complete, look for multiple complications at a time, and once a complication has been detected then the system should provide all the relevant detail related to that complication e.g. severity of the complication etc.

We have developed an EMR implementing case-based reasoning (CBR) for antenatal care with requiring very less typing and involve mostly the selection to facilitate the doctors to enter the records with ease. Once the record entering process begins the CBR module starts its working for predicting the diagnosis of the patient. CBR calculates the similarities (similarity scores) of the new case with the old cases and predicts the output of the new case based upon the output of previous similar cases. We used CBR for the prediction of the diagnosis in antenatal care because it has been used quite effectively especially when a patient has to be examined again and again [11]. Several authors have employed CBR in different areas of healthcare and have achieved significant success [11]-[15]. The results achieved were validated by cross validation technique and were found to be more than 93% accurate. 7% results were found to be false positive, so that is not a worry because no crucial instance was found to be reported as false negative. Our system was appreciated by the medical experts and the dataset prepared from different hospitals can also be used by scientific researchers to test their techniques.

The rest of the paper is organized as follows. The Background study has been presented in Section 2 followed by research methodology in Section 3. The results have been presented in Section 4 and a detailed discussion on these results is given in Section 5. Finally, Section 6 concludes the paper.

2 Related work

Case-based reasoning is a problem solving paradigm that in many respects is fundamentally different from other major AI approaches in the sense that instead of relying solely on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced, concrete problem situations (cases) [16]. CBR has been quite effectively used in medical domain for knowledge discovery. In [12] Brown et. al has discussed use of CBR in different medical projects. In medicine, CBR has mainly been applied for diagnostic and partly for therapeutic tasks [17]. In [17] Schmidt et. al have presented three solutions for adaptation of CBR. They suggest that Retrieval-only systems are especially useful for visualization tasks, because the users wish to see and interpret all specific details themselves [18]. In solving the adaptation by generalizing the condition holds that: Features get more typical as the case becomes more abstracted. This means the further down the cases are placed in the hierarchy, the more specific and less typical are their additional features in order to adapt by searching top down in a hierarchy of abstracted cases [17]. Another potential solution is Combining CBR with rule-based components. Some recent works [20], [21], [22] also used CBR for various medical and healthcare applications. In [23], authors present an optimization based approach, to enhance the quality of the tone-mapped LDR images using metaheuristics using differential evolution (DE) algorithm. Recently, an IoT based framework [24] proposed for COVID-19 patient management also showed the importance of managing the patient records. Similarly, authors in [25], [26], and [27] also worked on medical data for machine learning and image protection protection. However, to the best of our knowledge, application of CBR on antenatal care data is relatively new which presents a unique challenge because antenatal patients need to visit the physicians for follow-up visits several times during and after their pregnancy.

3 Methods

3.1 Dataset

The data consisted of real-world patients. The EMR developed for antenatal care was deployed in different hospitals. The data were carefully selected and entered so that we should not miss any vital feature. Some of the features were of Boolean type with values Yes and No, some features were numeric and rest were string type. The data were very complex because there are so many things which need to be noted and taken care off during antenatal care. There are also missing values in dataset which are replaced by using the method presented in [19]. The data also present a challenge to e-health researchers to check the performance of their proposed techniques.

3.2 Proposed system

Our proposed system serves two purposes for the medical practitioners and the scientific researchers. First, we have developed an Intelligent Electronic Medical Record (EMR) system for efficiency enhancement of antenatal care professionals. The EMR system has been developed keeping in view the fact that medical practitioners need quick and reliable information but they have very scarce time to interact with computers for this purpose. The system is designed in such a way that it facilitates doctors in storing all the relevant information of a patient while not requiring too much of the typing. We can say that our proposed technique offers the practitioners with flexibility to select that ask the query.

The second purpose of the system is to use case based reasoning (CBR) to predict the diagnosis of the patient. When a new case is entered in our system, old cases similar to new one are fetched from the database and then similarity score is calculated. These records are ranked according to their similarity score and subsequently, these ranks are stored in an array R. The top 5 elements of array R are selected and the records against those ranks are fetched to predict the diagnosis for the new case. The medical practitioner selects the appropriate choice of diagnosis to suggest appropriate treatment for the patient if it is found that any of those top ranked diagnosis correctly describes the state of the patient. At the same, the practitioner has the option to establish a new diagnosis for the patient if none of the top ranked diagnosis describes the state of patient correctly in their assessment. One advantage presented in our approach is that it presents the practitioner with all possible scenarios in the form of historical diagnosis so that we can make a much better analysis and treatment for mother and child. Secondly, automation presents doctors with all the factors that need to be considered while in manual approach, the doctors may overlook some important details due to any factor.

The security and integrity of the system has been assured by assigning only the registered users the rights to manipulate the data. In the presentation layer once the user logs in then a Graphical User Interface (GUI) provides the links to different modules of the system. The users under *admin* category can – through processing layer – insert new cases, retrieve old cases and also case based reasoning module helps them to predict the diagnosis of the patient. Processing layer contains two modules; i) Data entry system for antenatal care and ii) Case Based Reasoning (CBR) Module. The Data entry system module contains four sub-modules. The first sub-module collects the basic information about the patient like Patient ID, name, age, blood groups, education etc. and the previous treatment details – if any are also collected through this module. In the second sub-module the past history of the patient is collected. This sub-module includes past medical history, previous operations, drug history, family history and the data about previous pregnancies – if any. The third sub-module is the patient examinations module and includes Patient general examination, present pregnancy, Gynecological data, contraception details, conditional follow up details, Presenting complaints details, abdominal examination, per speculum inspection and vaginal examination details. Fourth sub-module collects the information about routine investigations and specific lab investigations. The assessment of the patient is also noted to validate the results obtained from the CBR module. The CBR calculates the similarities of the new case with already stored cases. The CBR module works as follows:

```
  //Initialize some variables
  SimScore = 0 //Similarity Score
  Nc = 0  // Number of attributes
  Initialize an Array R to store the rank of the records
  for all the t tables in the Antenatal database do the
following
    Store the newly added record in a dynamic array
    Nc = Number of Attributes of current table
    for i =1 to Nc
    Take ith attribute Ni of the newly added record
    Select ith attribute Oi from the corresponding table
from previous records – one at time
    if Ni = Oi
      SimScore = SimScore + 1/Nc
    End if
    Update the Rank array accrording to the SimScore
    End for
  End for
```

The results obtained through CBR were validated against the actual results specified by the doctors. The diagrammatic illustration of working of CBR module is illustrated in Figure 1.

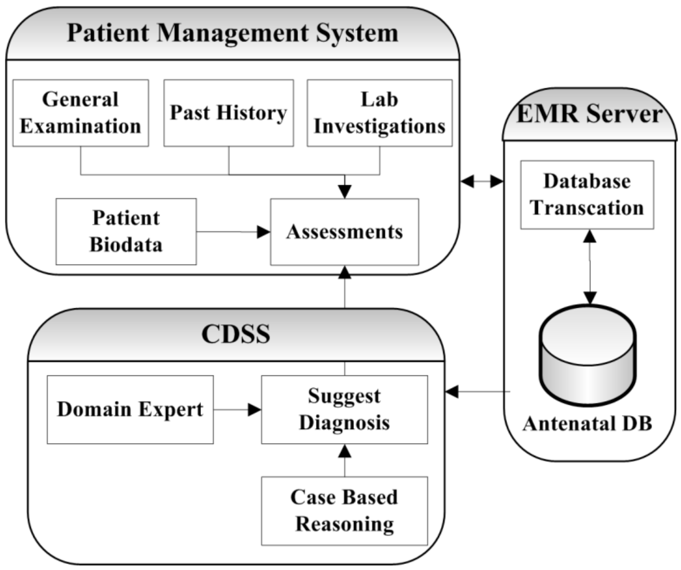

Fig. 1. System architecture

4 Results

The results are displayed in the form of a grid on the web form as given in Figure 2. The CBR module predicts existence of the following eight complications and their hierarchies – if any.

- Hypertension

  - PIH (Pregnancy Induced Hypertension)
  - Chronic
  - Pre-Eclampsia
  - Eclampsia
  - Hell Syndrome

- Diabetes
- Obstructed Labor
- Hemorrhage

  - Antipartum
  - Antipartum Placental Abruption
  - Antipartum Placenta Praevia
  - Post Partum

- Septicemia

  - Puerperal Sepsis
  - Septic Shock

- Septic Induced Abortions
- Thrombo Embolis

The CBR system displays the predictions – based upon the calculations by CBR – for above eight complications and their hierarchies and medical specialists have the choice to decide treatment for every complication accordingly.

4.1 Performance measure

We used cross validation technique to check the performance of our approach of prediction of diagnosis for antenatal care. Out of total 1200 records we randomly pick 200 records and calculate the similarity score by the method presented in Section III. The output predicted by the CBR is then compared to actual output

of the corresponding record. The performance is measured in terms of percentage accuracy as follows:

For Every complication and its hierarchy do the following steps:

$$\% \text{ accuracy} = \left(1 - \frac{FP}{TP}\right) * 100 \quad (1)$$

Where, FP is the number of false predictions, TP is the total number of predicted records. The percentage accuracy was to be 90%. Out of the remaining 10% records, 7% were false positive (FP) results, that is, our system predicted them positive for some particular disease but those patients were not suffering from that disease – as tested by doctor. These 7% FP did not harm at all because they were tested by the doctors for any disease and thus did not need any further treatment. 3% of the records were reported as false negative (FN) because their attributes were looking normal. The results are depicted in Table 1.

Table 1. Diagnosis Prediction for eight different complications

Simlarity Score	Hypertension	Diabetes	Obstructe d_Labour	Hamorhage	Septicaemia	Hamor hage	Septic_I nduced_ Abortins	Thrombo_ Embolism
18.34	NO	NO	NO	NO	NO	NO	NO	NO
18.23	NO	NO	NO	NO	NO	NO	NO	NO
18.17	NO	NO	NO	NO	NO	NO	NO	NO
17.82	NO	NO	NO	NO	NO	NO	NO	NO
17.81	NO	NO	NO	NO	NO	NO	NO	NO

5 Discussion on results

We have used CBR to predict diagnosis in order to assist doctors and scientists during different stages of antenatal care. The results obtained by CBR were compared with the original results. In most of the cases the results were accurate even in the presence of noise and missing data. In some cases the results were not accurate but most of the time the reported results were noted to be false positive which did not harm at all because the patients were reported positive for a complication which they did not had and their test was conducted as a precautionary measure. There were some cases when top 5 records contained both positive findings and negative findings for a particular disease. This issue was resolved by taking the majority votes. The medical specialists have the option to choose from the suggested diagnosis or he/she can ignore all the suggestions given by CBR as depicted in Figure 2.

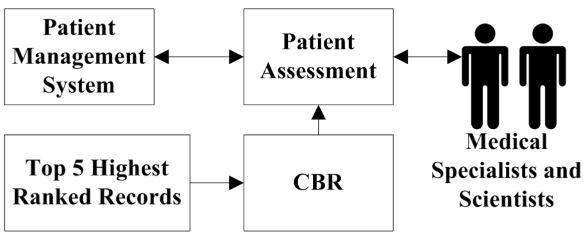

Fig. 2. The Medical Specialists and Scientists have the choice to make the final decision to choose from among the suggested diagnosis by CBR

We aggregated the result of all the patients according to mother condition, baby condition and delivery mode which are shown in Tables 2,3, and 4 respectfully. Class labels for mother condition are alive and dead. Class labels for baby condition are alive, dead and miscarriage. Finally, class labels for delivery mode are normal virginal delivery (NVD), lower segment cesarean section (LSCS) and evacuation of retained products of conception (ERPC). Mother and baby condition after delivery are comparatively shown in Figure 3. The CBR was found to be very effective while working with very complex and sparse data of antenatal care. The CBR also help in lowering the maternal mortality rate (MMR), miscarriage and baby death during pregnancy. The proposed CBR also helped to know about delivery modes as show in Figure 4.

Table 2. Condition of mother after delivery

Mother Condition	Frequency	Percentage
Alive	1149	95.75
Dead	51	4.25%
Total	1200	100%

Table 3. Condition of baby after delivery

Mother Condition	Frequency	Percentage
Alive	1172	97.66%
Miscarriage	3	0.23%
Dead	25	2.1%
Total	1200	100%

Table 4. Delivery mode

Mother Condition	Frequency	Percentage
NVD	940	78.33%
LSCS	258	21.50%
ERPC	2	0.17%
Total	1200	100%

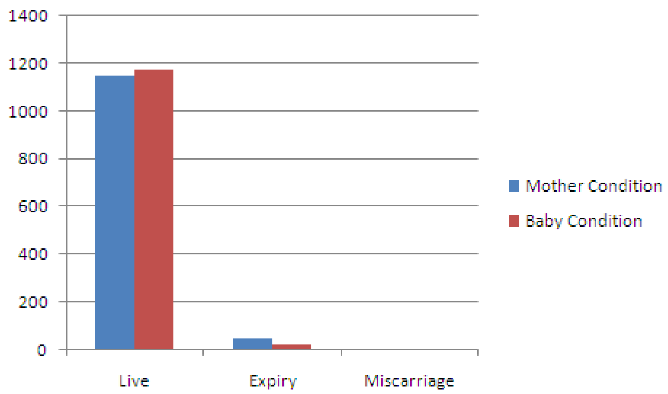

Fig. 3. Mother & baby condition after delivery

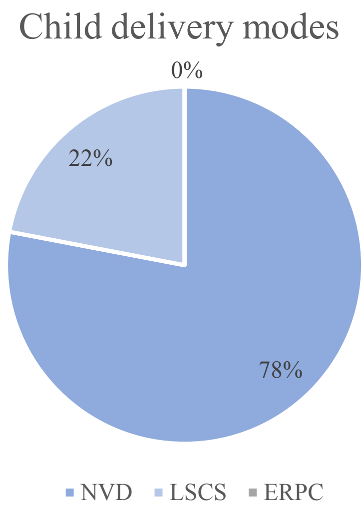

Fig. 4. Delivery mode

6 Conclusion

The paper presents a way to use ICT in the medical domain to improve health care specially e-health all over the world. Information technology concepts were applied to assist the medical practitioners to store the records of their patient in a database using a secure and easy to use environment and also to diagnose the patient. The Case-Based Reasoning was used to predict the diagnosis of eight major complications during the pregnancy for every new coming patient. Case-Based Reasoning was found to be very effective while working with sparse data like antenatal care where a patient generally visits the doctor very often to avoid any complication during and after the pregnancy. Future work will be focused on to select those attributes which actually effect every single complication. Another future direction may be to assign weights to some important attributes during the similarity score calculation process.

7 Acknowledgements

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-21-DR-73). The authors, therefore, acknowledge with thanks the University of Jeddah technical and financial support.

8 References

[1] Sutton, Reed T., David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. "An overview of clinical decision support systems: benefits, risks, and strategies for success." NPJ digital medicine, vol. 3, no. 1, pp. 1-10, 2020. https://doi.org/10.1038/s41746-020-0221-y

[2] Li, X., Wang, Y. and Li, D., "Medical data stream distribution pattern association rule mining algorithm based on density estimation", IEEE Access, vol. 7, pp.141319-141329, 2019. http://dx.doi.org/10.1109/ACCESS.2019.2943817

[3] Yoo Illhoi, Alafaireet Patricia, and Marinov Miroslav, Pena-Hernandez Keila, Gopidi Rajitha, Chang, Jia-Fu and Hua Lei, "Data mining in healthcare and biomedicine: a survey of the literature" Journal of medical systems, vol. 36, no. 4, pp. 2431-2448, 2012. http://dx.doi.org/0.1007/s10916-011-9710-5

[4] Zhang, Huiru, Guangshun Li, Yue Zhang, Keke Gai, and Meikang Qiu. "Blockchain-based privacy-preserving medical data sharing scheme using federated learning." In International Conference on Knowledge Science, Engineering and Management, pp. 634-646. 2021. http://dx.doi.org/10.1007/978-3-030-82153-1_52

[5] M. Kamran and M. Farooq, "An information-preserving watermarking scheme for right protection of emr systems," IEEE Transactions on Knowledge and Data Eng., Vol. 24, No.11, pp. 1950-1962, 2012. http://dx.doi.org/10.1109/TKDE.2011.223

[6] M. Kamran and M. Farooq, "A Formal Usability Constraints Model for Watermarking of Outsourced Datasets," IEEE Transactions on Information Forensics and Security, vol. 8, no. 6, pp. 1061-1072, 2013. http://dx.doi.org/10.1109/TIFS.2013.2259234

[7] Alemdar Hande and Ersoy Cem, "Wireless sensor networks for healthcare: A survey" Computer Networks, vo. 54, no. 15, pp. 2688-2710, 2010.

[8] Taghreed A, Lim SS, Sumi M, Bhutta ZA, Helga F, Matthews M, Jelka Z, Darmstadt GL. Achieving the millennium development goals for health: Cost effectiveness analysis of strategies for maternal and neonatal health in developing countries. In: British Medical Journal, 2005, pp. 1107. http://dx.doi.org/10.1136/bmj.331.7525.1107

[9] Khalid MZ, Akbar A, Tanwani AK, Tariq A, and Farooq M. Using Telemedicine as an Enabler for Antenatal Care in Pakistan. In: E-Medisys 2nd International Conference: Emedical System 2008.

[10] Mechel PN. Mobile phones for Mother and Child care (Case Study of Egypt): London School of Hygiene and Topcial Medicine, UK, May 2005.

[11] Bradburn C, Zeleznikow J. The application of case-based reasoning to the tasks of health care planning. In: S Wess, K D Altho, and M M Richter, Editors, Topics in Case-Based Reasoning. In: First European Workshop, EWCBR-93, Springer-Verlag Berlin, pp 1994. 365-378. http://dx.doi.org/10.1007/3-540-58330-0_100

[12] Brown, Daniel, Arantza Aldea, Rachel Harrison, Clare Martin, and Ian Bayley. "Temporal case-based reasoning for type 1 diabetes mellitus bolus insulin decision support." Artificial intelligence in medicine, pp. 28-42, 2018. https://doi.org/10.1016/j.artmed.2017.09.007

[13] Hsu CC, Ho CS. A new hybrid case-based architecture for medical diagnosis. In: Information Sciences—Informatics and Computer Science: An International Journal ed L. Medical Records, Medical Education and Patient Care. October 2004, pp. 231-247. https://doi.org/10.1016/j.ins.2003.11.009

[14] A SSR, Manickam S. Augmenting Medical Case Base Reasoning Systems with Clinical Knowledge Derived from Heterogeneous Electronic Patient Records. In: 14th IEEE Symposium on Computer-Based Medical Systems (CMBS'01), 2001, pp. 421-426. http://dx.doi.org/10.1109/CBMS.2001.941756

[15] Topel T, Neumann J, Hofestadt R. A medical case-based reasoning component for the rare metabolic diseases database RAMEDIS. In: Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), 2007, pp. 7-11. http://dx.doi.org/10.1109/CBMS.2007.9
[16] Aamodt A, Plaza E. Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. In: AI Communications, IOS Press, 1994, pp. 39-59. http://dx.doi.org/10.3233/AIC-1994-7104
[17] Schmidt R, Gierl L. Case-based Reasoning for Medical Knowledge-based Systems. In: International Journal of Medical Informatics, 2001, pp. 355-367. https://doi.org/10.1016/S1386-5056(01)00221-0
[18] Macura R, Macura K. MacRad: Radiology image resource with a case-based retrieval system. In: M. Veloso and A. Aamodt (eds.): Proceedings of 1st International Conference on CBR, Springer, Berlin, 1995, pp. 43-54. http://dx.doi.org/10.1007/3-540-60598-3_5
[19] A.K.Tanwani , Afridi, J., Shafiq, M.Z., Farooq, M.: Guidelines to select machine learning scheme for classifcation of biomedical datasets. In: EVOBIO 2009. LNCS, vol. 5483, pp. 128–139. Springer, Heidelberg (2009). http://dx.doi.org/10.1007/978-3-642-01184-9_12
[20] Lamy, Jean-Baptiste, Boomadevi Sekar, Gilles Guezennec, Jacques Bouaud, and Brigitte Séroussi. "Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach." Artificial intelligence in medicine., pp. 42-53, 2019. https://doi.org/10.1016/j.artmed.2019.01.001
[21] Chowdhury, Amrita Roy, Sreeparna Banerjee, and Tamojit Chatterjee. "A cybernetic systems approach to abnormality detection in retina images using case based reasoning." SN Applied Sciences, pp. 1-15, 2020. http://dx.doi.org/10.1007/s42452-020-3187-0
[22] Kouser, R. Ruhin, T. Manikandan, and V. Vinoth Kumar. "Heart disease prediction system using artificial neural network, radial basis function and case based reasoning." Journal of computational and theoretical nanoscience, pp. 2810-2817, 2018. https://doi.org/10.1166/jctn.2018.7543
[23] Almaghthawi, Ahmed, Farid Bourennani, and Ishtiaq Rasool Khan. "Differential Evolution-based Approach for Tone-Mapping of High Dynamic Range Images." International Journal of Advanced Computer Science and Applications 11 (2020): http://dx.doi.org/10.14569/IJACSA.2020.0110745
[24] Alam, F., Almaghthawi, A., Katib, I., Albeshri, A., & Mehmood, R. (2021). IResponse: An AI and IoT-enabled framework for autonomous COVID-19 pandemic management. Sustainability, 13(7), 3797. http://dx.doi.org/10.3390/su13073797
[25] Sugondo Hadiyoso, Farrel Fahrozi, Yuli Sun Hariyani, Mahmud Dwi Sulistyo, "Image Based ECG Signal Classification Using Convolutional Neural Network, International Journal of Online & Biomedical Engineering, pp. 64-77, 18(4), 2022. https://doi.org/10.3991/ijoe.v18i04.27923
[26] M Ravi Kumar,Kadoori Pooja, Meghana Udathu, J Lakshmi Prasanna, Chella Santhosh, "Detection of Depression Using Machine Learning Algorithms", International Journal of Online & Biomedical Engineering, pp. 155-162, 18(4), 2022. https://doi.org/10.3991/ijoe.v18i04.29051
[27] Mamtha Mohan, Suma K V, Deepika B Banagar, "An Efficient Framework to Protect Medical Images", International Journal of Online & Biomedical Engineering, pp. 143-154, 18(4), 2022. https://doi.org/10.3991/ijoe.v18i04.27841

9 Authors

Muhammad Kamran is working as Assistant Professor in Department of Cybersecurity, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. His research interest machine learning, data mining, cyber security, and data science.

Ahmed Abdul-Aziz Almaghthawi is working as Visiting Faculty Member in Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. His research interest machine learning, data mining, and data science.

Article submitted 2022-02-22. Resubmitted 2022-03-30. Final acceptance 2022-03-31. Final version published as submitted by the authors.