Recommendation System for Immunization Coverage and Monitoring

Uzair Aslam Bhatti[1,2,5], Mengxing Huang[1,2], Hao Wang[3], Yu Zhang[1,2], Anum Mehmood[4,6], and Wu Di[1,2]
[1]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, 570228, China
[2]College of Information Science & Technology, Hainan University
[3]Big Data Lab, Dept. of ICT and Natural Sciences, Norwegian University of Science and Technology, Aalesund, 6009, Norway
[4]Medical Officer, Abbasi Shaheed Hospital, Pakistan
[5]uzairaslambhatti@hotmail.com
[6]gooddranam@yahoo.com

***Abstract*—Immunization averts an expected 2 to 3 million deaths every year from diphtheria, tetanus, pertussis (whooping cough), and measles; however, an additional 1.5 million deaths could be avoided if vaccination coverage was improved worldwide.[1] New vaccination technologies provide earlier diagnoses, personalized treatments and a wide range of other benefits for both patients and health care professionals. Childhood diseases that were commonplace less than a generation ago have become rare because of vaccines. However, 100% vaccination coverage is still the target to avoid further mortality. Governments have launched special campaigns to create an awareness of vaccination. In this paper, we have focused on data mining algorithms for big data using a collaborative approach for vaccination datasets to resolve problems with planning vaccinations in children, stocking vaccines, and tracking and monitoring non-vaccinated children appropriately. Geographical mapping of vaccination records helps to tackle red zone areas where vaccination rates are poor, while green zone areas, where vaccination rates are good, can be monitored to enable health care staff to plan the administration of vaccines. Our recommendation algorithm assists in these processes by using deep data mining and by accessing records of other hospitals to highlight locations with lower rates of vaccination. The overall performance of the model is good. The model has been implemented in hospitals to control vaccination across the coverage area.**

***Index Terms*—Health Recommendation System, Health Information System, Decision Support System, Big Data for Health Analysis**

I. Introduction

Vaccination has greatly contributed to eliminating the burden of various infectious diseases. Vaccination reduces the spread of disease in children. The recommended immunization schedule is planned to protect infants and children early in life when they are most susceptible and before they are exposed to potentially life-threatening diseases, as shown in Fig. 1. Some vaccinations are administered at the time of birth, while the rest are after given months later, following a schedule. It is highly recommended that the vaccination schedule be followed to positively impact the health of each child. However, complex issues have arisen regarding vaccination planning, as many unaware parents do not follow the vaccination schedule; therefore, it is difficult for the vaccinator, the health care staff member who administers vaccines, to administer the correct vaccinations at the appropriate times. For example, if the Measles 1 vaccine is not given at the right time, then the duration between Measles 1 and Measles 2 varies because a particular gap is recommended. Our team completed surveys of hospitals in Pakistan and identified many gaps in the understanding and implementation of vaccination planning. In general practices, a vaccination card is provided to parents for each child's vaccination plan, and they are advised to bring this card each time the child comes in for vaccination. However, due to low literacy rates and a lack of awareness, parents often lose the card and sometimes do not remember the vaccination name or date.

Vaccine	Birth	1 mo	2 mos	4 mos	6 mos	9 mos	12 mos	15 mos	18 mos
Hepatitis B[1] (HepB)	1st dose	←2nd dose→			←3rd dose→				
Rotavirus[2] (RV) RV1 (2-dose series); RV5 (3-dose series)			1st dose	2nd dose	See footnote 2				
Diphtheria, tetanus, & acellular pertussis[3] (DTaP: <7 yrs)			1st dose	2nd dose	3rd dose			←4th dose→	
Haemophilus influenzae type b[4] (Hib)			1st dose	2nd dose	See footnote 4		←3rd or 4th dose, See footnote 4→		
Pneumococcal conjugate[5] (PCV13)			1st dose	2nd dose	3rd dose		←4th dose→		
Inactivated poliovirus[6] (IPV: <18 yrs)			1st dose	2nd dose	←3rd dose→				
Influenza[7] (IIV)					Annual vaccination (IIV) 1 o				
Measles, mumps, rubella[8] (MMR)					See footnote 8		←1st dose→		
Varicella[9] (VAR)							←1st dose→		
Hepatitis A[10] (HepA)							←2-dose series, See footnote		

Fig. 1. Vaccination schedule for children under 18 months[2]

Our approach in this paper is to automate this system for hospitals, vaccinators and children, allowing a comprehensive system to solve these immunization problems. The process

[1] Data source for immunization records of 1.5 M: http://www.who.int/mediacentre/factsheets/fs378/en/

[2] Data for the vaccination schedule have been taken from the CDC website and are available online. https://www.cdc.gov/vaccines/schedules/downloads/child/0-18yrs-child-combined-schedule.pdf

begins with the registration of the child and ends with the recommendation of a vaccine. The proportion of children [1] aged 12–23 months who are fully vaccinated by 12 months of age has been decreasing; this proportion was 67% in 1992, then 55%, 54% and 51% in 1996, 2000 and 2004, respectively. Analyses have also shown that the birth order of the child, residence (rural/ urban) and mother's education are major determinants of the immunization status of the child. Delays in childhood vaccines [2] have increased in recent years. We focus on geographic clusters of under-immunization and vaccine refusal, then compare clusters of under-immunization for various types of vaccines, and finally evaluate whether vaccine refusal clusters pose barriers to achieving high immunization rates. Vaccination [3] helps economic growth everywhere because of lower morbidity and mortality. The annual return on investment in vaccination has been calculated as between 12% and 18%. Vaccination also leads to increased life expectancy. Long healthy lives are now recognized as a prerequisite for economic growth on personal and national levels, and economy promotes health. Vaccines are thus efficient tools to reduce disparities in economic conditions and inequities in health.

Immunization saves millions of lives by preventing disease. Through financial support from many non-government organizations such as the GAVI Alliance and, more recently, the Bill & Melinda Gates Foundation, significant advancements have been made in immunizing children since 2000 via the Global Immunization Vision Strategy (GIVS). The child mortality rate has decreased in the last few years, but the greatest challenge is due to daily increases in birth rates, which has drawn significant attention from governments. In one survey, it was reported that 24 million children[3] born every year do not receive proper immunization during their first year of life.

Immunization is a highly[4] economical tool when everyone is required to participate. Immunization curricula currently reach over approximately 80% of the world's children, which helps reduce the death ratio by preventing over 2 million deaths annually. In May 2012, the World Health Assembly approved the Decade of Vaccines Global Vaccine Action Plan (GVAP). A guiding principle of the GVAP is country ownership of national immunization programs, one measure of which is the proportion of domestically financed program costs. By 2020, the GVAP aims for all countries to be properly preparing and managing their immunization budgets, with their immunization programs sustainably financed.

[3] Katib, Anas, et al. "Jeev: a low-cost cell phone application for tracking the vaccination coverage of children in rural communities."

[4]Oxford University Press in association with The London School of Hygiene and Tropical Medicine https://academic.oup.com/heapol/article/30/3/281/614291/An-analysis-of-government-immunization-program

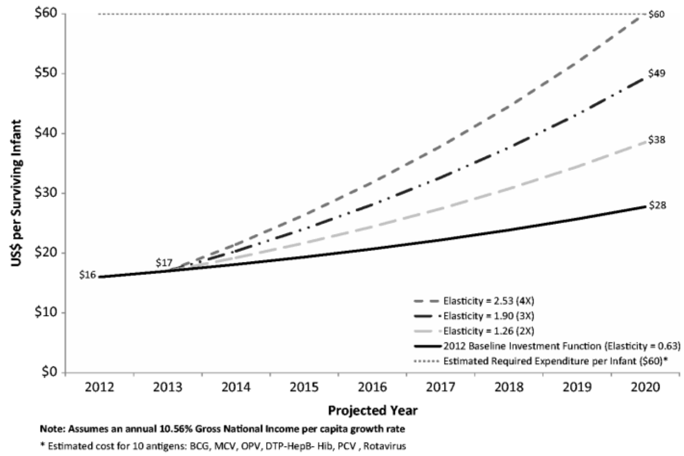

Fig. 2. Projected governmental routine immunization expenditures per surviving infant under varying immunization investment functions, based on reporting GAVI-eligible countries in 2012

II. Related work

An immunization information system (IIS) is an important global priority, and online access to that information is crucial. In [4], the effect of online access to state immunization practices for emergency department (ED) providers was estimated for a pediatric academic tertiary care center. Another approach [5] to reduce immunization problems worldwide is education; it is believed that public health efforts to address problems of vaccine hesitancy should increase their focus on childhood education. One strategy to address the issue of vaccination is to include a fun aspect; it is believed that by having fun, children can be educated about the need for vaccination. The current digital world can help address this issue in multiple ways that complement immune system and immunization education in school curricula.

Management information system (MIS)-based approaches [6] can use smart phone applications. One software application, called Jeev, is helpful for monitoring the immunization issue across India. The focus of Jeev was to combine the unique ID information of children with immunization information. Initially, the project was focused on rural areas to assess the sustainability of the project. In [7], the immunization of healthy nodes, in the presence of already infected nodes, was studied. A robust algorithm was proposed to identify and resolve such a problem and to help health professionals make more informed choices and tailor their decisions to the actual distribution of the epidemic on the ground. Data-aware vaccination problems were highlighted by the author, who demonstrated that these problems are hard to approximate. Three effective polynomial-time heuristics, DAVA, DAVA-prune and DAVA-fast, of varying degrees of effectiveness and reproducibility have been proposed. Finally, the scalability and efficacy of algorithms have been revealed through extensive experiments on multiple real networks, including large epidemiology datasets. In [8], the authors consider the problem of rapid spread of an epidemic in an arbitrary contact network by issuing vaccination status throughout the network. The authors consider the problem of vaccine distribution across the network, as shown in Fig. 3, and they track the issue to find an optimal cost distribution of vaccination resources where different levels of vaccination are allowed.

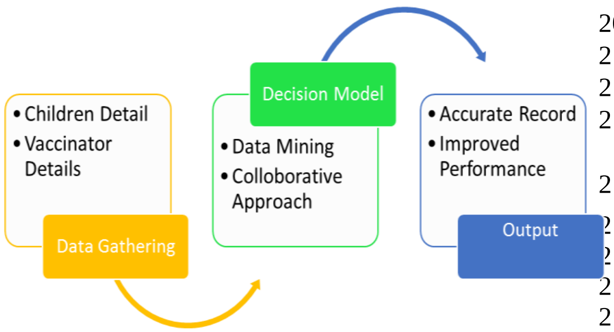

Fig. 3. Immunization data collection phases for analysis

Another method [9] attempts to remove this assumption from vaccination uptake estimation and proposes an algorithm to classify data from online clinical records and other sources. Comparisons of the data with web data are performed using state-of-the-art techniques. The performance improvement is more notable for vaccines that have been implemented more poorly due to media attention that is negative (HPV-1 and HPV-2), problems in the vaccine supply (DiTeKiPol), or the targeting of children 12 years old (whose vaccinations are more irregular compared to younger children). Another approach [10] explains that the major goals of most national, international and non-governmental health organizations are to eradicate the occurrence of childhood diseases that are preventable by vaccine (e.g., polio). Pakistan has been trying to eliminate polio for many years but has not been successful; each year, some cases are identified in rural areas, where the education level is poor. It is important to realize that children should receive proper vaccinations at least at the starting 2 years because so many vaccinations should be administered during this period of time to help children grow up disease-free and to help the government facilitate a healthy generation. In the rural areas of many countries, there are no tools to identify individuals and no documentation to track the vaccination schedule of a child; therefore, in another paper [10], the authors focused on mapping the finger prints of children and parents for use as a unique ID of the vaccinated child so that the child's vaccination schedule can be later identified without any documents. Fingerprints can be effectively used to recognize children from birth to 4 years. A total of 1,600 fingerprint images (500 ppi) of 20 infants and toddlers were captured over a 30-day period. Various algorithms for a biometric recording of the children were then performed so that the accuracy could be improved.

A cold chain information system (CCIS) [11], which tracks health facility information and a country's cold chain equipment, can help address these challenges at both the strategic (planning) and tactical (management) levels. In this report, the authors describe the work needed to develop a CCIS for Laos, in alliance with UNICEF and the Lao Ministry of Health. This work is based on experience in deploying CCISs in multiple countries, along with the development of analysis tools and data standards. The technical contributions of this work include the development of a new data model for DHIS2 and a text message system using an Android gateway to RapidSMS with a custom notification engine interfacing to DHIS2.

III. Motivation

Government funding for immunization in many countries has increased in recent years and is complemented by donor funds, as discussed earlier. However, equitable and universal access to vaccine and immunization services in all countries is still required. Different countries have their own methods for monitoring and implementing immunization programs. We have proposed a model that helps to improve decision-making in vaccination strategies across a province or a country. In addition to developing applications [3] to register and monitor children, a tool is needed that can comprehensively suggest government/NGO strategies for current vaccinations in an area, while also identifying the exact number of children in an area that are currently unvaccinated or need to be vaccinated in the near future. Vaccinators used in this project for remote area monitoring send data to the server from a mobile application developed in-house.

We used an open source data toolkit named Open Data Kit (ODK) to develop the mobile forms, as this kit does not require proficiency in mobile app development; AngularJS was used to obtain analyzed data from the server. The data mining algorithm is run on a centralized server, in the same location that the hospital management system is integrated, so patients can be mapped with children as two types of data, as shown in Fig. 4. In addition, to improve the analysis, we used a dataset from the current hospital and a dataset from another hospital, which were provided by the government and other NGOs; future global approaches will further enhance the algorithm.

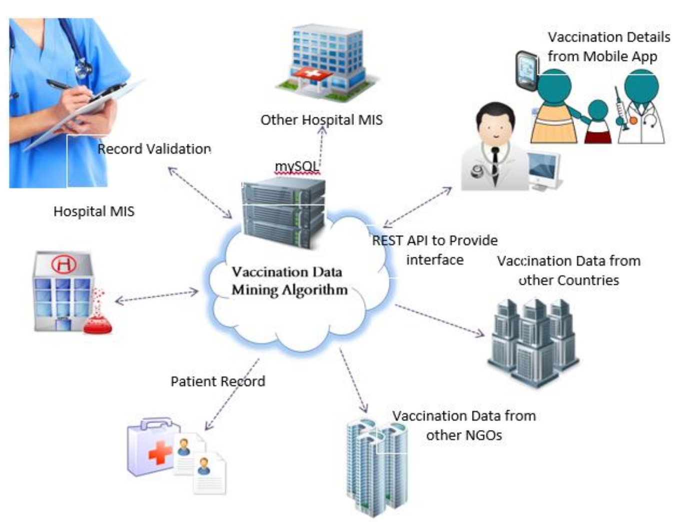

Fig. 4. Data mining model to improve immunization problems

Our recommendation model also suggests areas for mobile vaccinators to visit and the vaccines that should be supplied. The algorithm helps plan for appropriate stocks of vaccines so that sufficient vaccines are available. Children under the age of 2 years are the targeted population of our model. The working principle of the model is based on a collaborative approach using data mining to show effective ways to vaccinate areas with low immunization rates.

For vaccine recommendations for children, we implemented a data mining approach with a collaboration that helps suggest the correct vaccine based on historical data from vaccination centers in which the vaccines are stored centrally, where C_a - represents the child's age, V_i is the vaccine identified, C_i represents the child's unique ID, d is the dataset for a given hospital, and d_o represents data for another hospital.

Algorithm 1 (needs better explanation in the text)
Require: Input C_a, V_i, C_i, d and d_o 1: Sample S_d from d and d_o 2: Select the nearest nodes from d and d_o (KNN approach) 3: Build the tree to select the nearest age child N_k or select the nearest schedule from record R_v 4: $V_s = \emptyset$ (vaccination selected, default zero) 5: for i ⟵ 1 to N_k 6: Match C_a = d or d_o 7: If C_a not found, 8: Match $C_a = R_v$ 9: End if 10: $C_a = \{d \text{ or } d_0\} \cup \{R_v\}$ 11: End for 12: Return C_i

The above algorithm selects the child's record from historical data and compares it with the standard schedule. If the record is not available, then the best matching case is applied based on the child's age in the previous record. The above procedure helps identify how many children are not vaccinated and have deficiencies with regard to particular vaccines. We applied the above method to a dataset from the hospital and found many records, which aided in the tracking of vaccination records with missing immunizations and in reducing disease occurrence in children.

Our decision support system informs vaccinators as to which area will have a high priority for vaccination and how many vaccines are required in the future: if V_i is the vaccine ID, V_N is the non-vaccinated child, and C_i represents the total number of children, then the total average N of unvaccinated children will be

$$V_N = \sum_{k=0}^{n} \frac{(C_{ik} + V_{ik})}{N}, (1)$$

Our model is helpful for caregivers and can track the schedule of vaccinations in several ways. Figure X demonstrates the total number of vaccinations required of vaccinators in the next month. The dataset collected from the hospital was analyzed in different ways, as shown in Fig. 5.

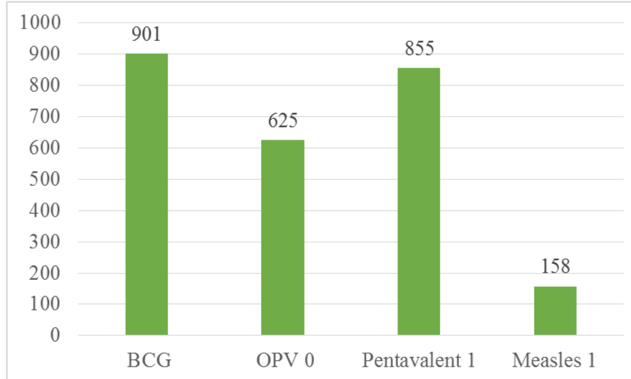

Fig. 5. Immunization trend over the previous month in a hospital

This analysis is helpful for vaccination schedules and for evaluating vaccine stocks because it calculates the average vaccine consumption within the last month. A percentage increase can also be evaluated from that trend; therefore, decision-making for the organization is simplified.

Table 1 shows the total number of children vaccinated by the organization within two months, which helps to predict the percentage increase in vaccinations so that the required vaccinations are available.

Table 1: Immunization in 2017 in a hospital

S.No	Vaccinations	Jan	Feb
1	Bcg	1727	2727
2	Opv-0	1522	1422
3	Opv-1	1153	1153
4	Opv-2	806	852
5	Opv-3	845	1165
6	Pentavalent-1	1165	2110
7	Pentavalent-2	806	844
8	Pentavalent-3	852	971
9	Pneumococcal-1	1165	1754
10	Pneumococcal-2	809	852
11	Pneumococcal-3	844	1165
12	Ipv	818	809
13	Measles-1	613	684

Our decision support system can predict stock availability from Eq. (2), where S_v is the stock of vaccines, V_A is the average consumption of a particular vaccine over the last few months, and D_n is the average dropout of vaccines, that is, the vaccines lost over the last few months:

$$S_v = \left(\frac{\sum_{n=0}^{N} Vn}{N}\right) + \left(\frac{\sum_{n=0}^{N} Dn}{N}\right). (2)$$

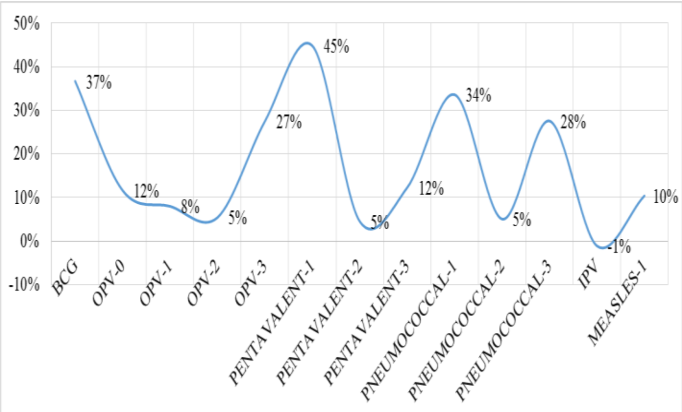

Fig. 6. Percentage increase of vaccinations required

We also produced Fig. 7 using this vaccination module with a health management system so that trends in diseases and causes of diseases can be monitored. Our approach is comprehensive and helped the hospital identify issues and make decisions.

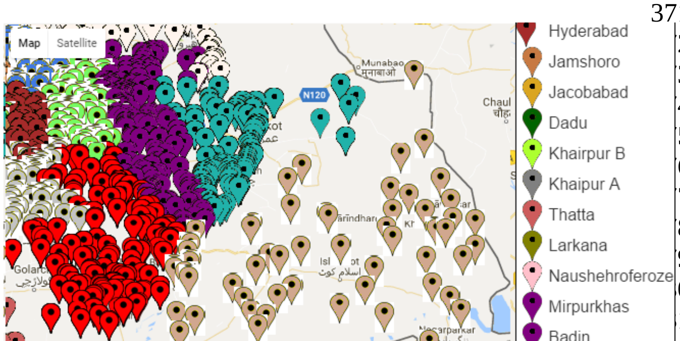

Fig. 7. Graphical representation of vaccination zones

The map can help one analyze important areas for management decision-making. Thus, if governments are making large investments (up to 20$ per surviving infant in 2020, as shown in Fig. 2) in immunizations, they should implement an effective monitoring framework to improve performance. Our study is limited to monitoring the immunization of children less than 2 years of age because of the increased mortality rate in this age range.

IV. Conclusion

In this paper, we analyzed our results using a hospital database in Pakistan and showed that it is possible to track issues with vaccination through our data mining model. A detailed study is needed to scale up this model across different locations, and a unique collective approach can help us to solve the problem of immunization across the world. Maps are also helpful for tracking red zone areas and highlighting key problem areas. Significant daily changes are currently being observed in the immunization programs of hospitals. Current work is focused on evaluating the use of IoT devices [what is this device]; the use of these devices is growing for medical purposes; this work will automate a 100% unattended immunization model and, in the future, will have the ability to be replicated anywhere. Moreover, the distributed vaccination model should also be improved for the vast areas in which internet connectivity is a serious issue; otherwise, the recommendation results will not be beneficial for people in those areas.

References

[1] Munthali, Alister C. "Determinants of vaccination coverage in Malawi: evidence from the demographic and health surveys." *Malawi Medical Journal* 19.2 (2007): 79-82.

[2] Lieu, Tracy A., G. Thomas Ray, Nicola P. Klein, Cindy Chung, and Martin Kulldorff. "Geographic clusters in under immunization and vaccine refusal." *Pediatrics* 135, no. 2 (2015): 280-289.

[3] Andre, Francis E., et al. "Vaccination greatly reduces disease, disability, death and inequity worldwide." *Bulletin of the World Health Organization* 86.2 (2008): 140-146.

[4] P. R. Shankar, Michael Scott Kelleman, C. E. McCracken, C. R. Morris and H. K. Simon, "Real time access to online immunization records and its impact on tetanus immunization coverage in the ED," *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*, Mysore, 2016, pp. 1-6.

[5] Kumanan Wilson, Katherine Atkinson & Natasha Crowcroft (2017) Teaching children about immunization in a digital age, Human Vaccines & Immunotherapeutic, 13:5, 1155-1157.

[6] Katib, Anas, et al. "Jeev: a low-cost cell phone application for tracking the vaccination coverage of children in rural communities." *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*. IEEE, 2013.

[7] Zhang Y, Prakash BA. Data-aware vaccine allocation over large networks. ACM Trans Knowledge Discover Data. 2015;10(2):20:1–32.

[8] Preciado, Victor M., et al. "Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks." *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013.

[9] Dalum Hansen, Niels, et al. "Time-Series Adaptive Estimation of Vaccination Uptake Using Web Search Queries." *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017.

[10] Jain, Anil K., Kai Cao, and Sunpreet S. Arora. "Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage." *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. IEEE, 2014.

[11] Anderson, Richard, Trevor Perrier, Fahad Pervaiz, Norasingh Sisouveth, Bharath Kumar, Sompasong Phongphila, Ataur Rahman, Ranjit Dhiman, and Sophie Newland. "Supporting immunization programs with improved vaccine cold chain information systems." In *Global Humanitarian Technology Conference (GHTC), 2014 IEEE*, pp. 215-222. IEEE, 2014.