Case Report

Data Engineering Frameworks for Optimizing Community Health Surveillance Systems

Venkata Krishna Azith Teja Ganti [1*]

[1] Support Engineer, Microsoft Corporation, Charlotte NC, USA

*Correspondence: Venkata Krishna Azith Teja Ganti (venkatakrishna.azith.de@gmail.com)

Abstract: A Changing World Demands Optimized Health Surveillance Systems – and How Data Engineering Can Help There is a growing urgency to manage the public health and emergency response practices effectively today, in light of complex and emerging health threats. Fortunately, a host of new tools, including big and streaming data sources, methods such as machine learning, new types of hardware like blockchain or secure enclaves, and means of data storage and retrieval, have emerged. But, with these innovations comes a grand challenge: how to blend with, and adapt them to, the traditional public health practices. The long-in-place infrastructures and protocols to protect and ensure the welfare of communities are in need of change, or at least update, to enhance their marked longevity of impact directly on the health outcomes and community wellbeing they were designed to fortify. It is in this vein that the essay is written and composed. The investigation in this essay is to query what, particularly, might be the aspects and influences of the emerging veritable cornucopia of new data engineering frameworks that are either being developed specifically for health surveillance and wellness, or are available to be co opted from devices and services already thriving in the current market and research milieu. Knowing what these ways may be could well aid in molding their uptake and spread, ensuring their beneficial impacts on those communities who stand to gain the most. The essay is divided into several key segments. After this introduction, section two details the research methods. In the section that follows, the maximum health outcome potentials of these novel frameworks are reviewed. Part four of the essay takes a more critical approach, addressing how the success of these methods may be hindered and future research avenues. Lastly, the concluding information suggests some actions to take to aid best suit the implementation of these ways, and suggests some thoughts for further research after the completion of these inquiriestrand [1].

Keywords: Data Engineering, Public Health, Surveillance Systems, Community Health, Health Data Integration, Community Health Analytics, Surveillance System Optimization, Real-Time Health Monitoring, Data Pipeline Automation, Predictive Health Modeling, Public Health Data Management, Health Information Interoperability, Big Data in Epidemiology, Health Surveillance Data Framework

1. Introduction

Community health surveillance has become a critical approach for preventing, controlling, and mitigating public health challenges – for example, disease outbreaks, natural disasters, and bioterrorism – in villages, schools, and mosques. With advances in both technology and healthcare, digital and information data science has been extensively applied in public health. Data analytics could provide meaningful knowledge for clinical doctors, public health officials, and health-related patients . Properly mined information and data could also effectively help in decision making for management and strategy in healthcare services. Undoubtedly, community health surveillance systems are important.

How to improve the effectiveness of community health surveillance has become a crucial issue for health and medical informatics researchers. In the classic literature of the health effect model, traditional tools and techniques are usually used to explore this issue. Research, such as experiments or observation, is designed to consider the health outcome as an input variable in community health surveillance systems. As the quantity of subsequent social and health-related data grows, how to handle or mine data is an effective way to deal with health and medical informatics . In the development of data engineering, some predictive models can be used to analyze health data or health outcomes produced by data radar frameworks. As a result, the health information produced daily is directly changed to generate clinically meaningful knowledge. The research raises some concerns about the techniques of data-capture radar frameworks when applied to health data in community health surveillance systems. A series of secondary questions are raised to develop these issues. The research aims to further these issues. Finally, based on the assessment of community health surveillance systems, this study conducted illustrative analytical experiments.
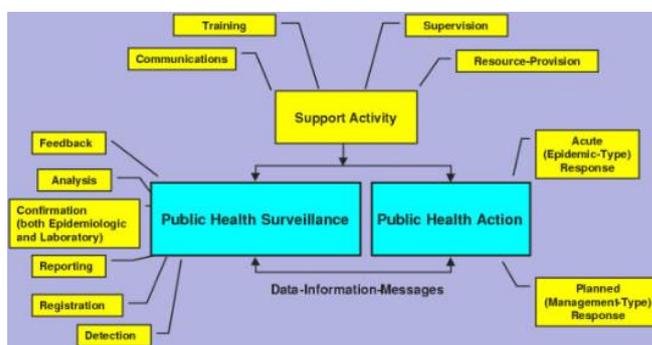

Figure 1. Conceptual framework of public health surveillance and action

1.1. Background and Significance

This year marks the 40th anniversary of the first Community Health Monitoring System, and the 20th anniversary of the HealthSTAT Community Health Monitoring System. These computer-assisted systems for printing monthly summary Public Health Reports facilitate an up-to-the-minute analysis of community-level health data including hospital, clinic and ambulance data, and vital records. Theoretically, a computer, a printer and a telephone line are all that is needed to operate such systems, and the HealthSTAT program has earned this reputation. HealthSTAT or a variation now operates in 49 states in 145 public and 30 private health departments, and in 26 other countries. In the United States new interest and activity in community health statistics and community health status indicators has been greatly stimulated by HealthSTAT and its spin-offs. As funds become available, new community health monitoring and analysis systems will be implemented and existing hospital-based monitoring systems will be shared with HealthSTAT-like systems in state and city health departments. Similarly, HealthSTAT monitoring and analysis centers will be implemented in states where hard hit areas do not now have a center. For a while the HealthSTAT "system" then may become a national health monitoring and analysis "network", where some health departments function as producers of local data or indicators and others function as consumers of previously compiled or analyzed data. To assist readiness of monitoring and analysis software for these activities, an improved version of the HealthSTAT software package has been developed. This software is designed for IBM PC compatibles. It is menu-driven with context-sensitive help screens and is easy to use. It also has a number of state-of-the-art features including an English Macro language for users who want to write more complicated programs. It can interface to many other database management systems.

Equation 1: Decision Support System (DSS) Optimization

$$DSS_{opt}(t) = \max\left(\sum_{l=1}^{q} W_l \cdot Y_l(t)\right)$$

where:

$DSS_{opt}(t)$ is the optimal decision support output at time t,

$Y_l(t)$ is the relevant decision-making factor l at time t,

W_l is the weight or priority of each factor,

q is the number of factors involved in decision-making.

1.2. Research Objectives

To have an effective community health surveillance system, good data engineering frameworks are needed. It is important to identify existing data engineering frameworks that are currently used or can be applied to community health surveillance systems, and to explore how well they perform or what improvements may be needed. This research has the objective to investigate the data engineering frameworks for community health surveillance systems. The research questions driving this exploration include: How can data engineering frameworks be used and what is available in the context of community health surveillance systems? Are those data engineering frameworks appropriately used and effective in community health surveillance systems? What are the performance or other data needs of data engineering frameworks in community health surveillance systems? Based on research findings, what data engineering frameworks are needed at a community health surveillance system and how should it be developed / improved? Finally, what is the potential impact of data engineering frameworks developing and improving in community health surveillance systems? The concept of community health surveillance covers syndromic surveillance, event detection, and situational awareness. These involve monitoring relatively small spatial environments consisting of communities, such as a single event, or monitoring the health of relatively large populations consisting of complex systems . The common goals are to monitor public health in near-real time in order to quickly detect new health threats or health system disruptions. There have been a few community health systems designed, focusing mainly on hand-held or wireless devices, and there have been no reports found that were specifically internal systems used for health surveillance. Moreover, those that do cover general public health needs are not getting into very specific internal health issues like those of a community health system.

2. Community Health Surveillance Systems

The rapid increase of globalization and human activity complexity has affected the world's ecological environment especially in densely populated urban areas. One of the sectors that have been affected is the traditional public health sector, which is currently termed for community health. As public accessibility gets wider, the amount of data within every healthcare institution including public health institutions plays a significant role in the fulfillment of community health information. This can potentially offer a great opportunity to build a comprehensive data-driven community health surveillance system by utilizing collected healthcare data. Community health surveillance systems include data acquisition, storage, analysis, use, and feedback. The performance in community health surveillance systems relies on components of surveillance systems. The performance of the surveillance system decreases if one or more of these components fail .

Community health surveillance systems are employed for monitoring and analyzing trends in public health. Accumulated data from these surveillance systems is used by

health officials to formulate and monitor policies for disease prevention and health promotion. It is common practice to join elementary reports from multiple sources when performing disease surveillance. There are four usual types of data that contribute towards health surveillance. They are data on the health status, collected either for individual cases or for the population at large, data on risk factors for health, data on health service treatment provision or consumption, and data on different health producing determinants such as environmental factors and lifestyle. The use of these data is two-fold: to understand the health situation better and to advocate changes in policy that combat the emerging public health challenges.

Community Health Surveillance for public health institutions is an essential activity for monitoring public health trends and possible disease outbreaks. Event or case based systems are employed for the collection of data reflective of pre-specified sources. The collected data is then assessed against pre-given epidemiological criteria. The effect of this assessment is expected to be a public health action like an investigation, reaction or feedback. Monitoring is a public health function that is frequently in evidence where conditions permit. Analysis and interpretation of routine data is essential to the public health response to surveillance. Plurality of action is possible at the analytical stage. There is a direct relation between data and public health action. Data should be sufficient to prompt a clear understanding of the root problem and the mitigation strategy. Surveillance data is collected from a variety of sources. Data must also be valid. The type of action is also considered. Prompt data is best suited to immediate containment or prevention measures while a more thoughtful complex issue requires complex analysis. As a result, benchmarks are essential, most often in the form of statistical norms. Fortunately, the amplitude of quantitative data gathered by public health institutions is very large. Public health institutions organize on community health policy but they frequently have to collect data from all public health institutions providing healthcare services. Some other morbidity data are supplied directly but they cannot be considered exclusive from an official authority and may contain inaccuracies. In order to interpret the data, the epidemiological capacity of public health institutions needs to be utilized. Public health institutions, instead, are involved in the collection and analysis of data but resources are very limited, thus effectiveness is in question. There are also other reasons for problematic analysis of morbidity data within public health authorities. Data transmission is not entirely efficient because of paper-based systems. Furthermore, some districts do not inform public health institutions well enough because they are at a greater focus on their performance assessment. Data is also not transmitted accurately; this is thought there are efforts to prevent this. Participatory data management is intended for users to participate in data management through editing and validation on sighting. There are many indications which show the quality of data within the system is deficient. The timeliness and accuracy represent the biggest problem. Episodes are frequently reported many days after the occurrence, diagnosis of episode is often incorrect.

Figure 2. Surveillance systems

2.1. Definition and Components

[2] Public health surveillance has long been an essential public health strategy to monitor, detect, investigate, and respond to threatening health events. The design and

implementation of community health surveillance systems are informed by the conceptual model that characterizes the systems with their vision, mission, objectives, components, key strategies, and bi-directional relationship with technology enhancement . One key strategy is the activities that support data collection, processing, analysis, and dissemination. Herein, the term "public health surveillance activities" is adopted to describe these essential components when they are performed together in a systematic way to achieve high relevance and cost-effectiveness. The four public health surveillance activities comprise community health surveillance systems – direct observations, open data monitoring, periodic health summits, and mobile health or telehealth practices. They are illustrated by various examples of surveillance activities in low-resource settings to better demonstrate how the theoretical framework can be applied in practice. In describing or designing a public health surveillance activity, effective interrelationship between data collection, processing, analysis, and dissemination is necessary to maximize the outcomes of a surveillance system. The importance of technology and data engineering will be greatly addressed to understand the complexity of current and future public health surveillance systems, and how data and technology could be effectively engineered to optimize the overall performance of the systems. Amid rapid evolution in technology and data science, the overall design and operation of current health surveillance systems also tend to become more complex and dynamic. A strategy for designing effective and efficient community health surveillance systems is greatly needed for health program managers, designers, and policy-makers to optimize, up-scale, and maintain national health surveillance systems. This focuses on technology and data engineering frameworks for public health surveillance in community settings.
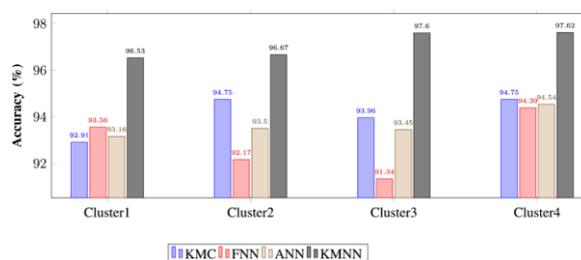

Figure 2. Edge computing based secure health monitoring framework for electronic healthcare system

2.2. Challenges and Limitations

Community health surveillance systems today aim to assist decision-making on community health issues by providing real-time or near real-time information. However, compared to vital statistics or research data, the practicality of the surveillance data that are currently used is questioned because of its limited range. Real-time health surveillance systems require various other health data, including routinely recorded electronic health records (EHRs) in hospitals, health centers and pharmacies. Data sharing among various organizations is essential for a successful health surveillance system; however, the current health data in such agencies are mostly non-public knowledge. Given this situation, it is now discovered that a well-designed health surveillance system is particularly helpful for community health monitoring only when a similar accident or pandemic occurs in a confined area . Resource distribution in the network, which is planned against ten types of pests and catalogs, will malfunction if several losses are connected to one control-node, and a node of event-water-borne disease occurs to other areas.

Instead of directly making decisions, health systems can influence various actors who have authority in the field of public health in different ways— by arming them with information that supports a claim needs action, by increasing public awareness of a situation to catalyze action from the public or politicians, and by providing guidance or

norms on appropriate responses. Political choices shape outcomes in health, and health surveillance can be a means by which political choices are informed. But the process of a government becoming concerned to public release and response following the passage of the Public Health Act can be opaque. Ensuring that transparency and accountability and public disclosures through an elected body, even in the presence of special advisers and necessity, needs to be respected. Amplifying uncertain or incorrect information during a health crisis can inappropriately cause unnecessary public alarm. The temporal effects during paroxysms of fever and panic of those witnessing the suffering of those already infected were deployed by campaigners in the First and Second World Wars to develop propaganda to engender sympathy within neutral countries. If the tipping-point in existing host-resilient control using telemedicine strategies is crossed, the rate of further increase in transmission until population recovery will increase. In this context, information gaps are rife, causing ethical dilemmas. For instance, a non-observant pertussis infection in the domains at its tipping point of becoming a major epidemic in humans.

3. Data Engineering in Public Health

Understanding the context of data collection, and the specific uses of that data for decision support, is a prerequisite for solid public and population health informatics. Public health frameworks for the collection, analysis, and dissemination of such data need to be understood by those who do systematic reviews. The decision must also be made about when data-related topics overlap with public health informatics fields such as eHealth, mHealth, and consumer health informatics. Therefore, knowledge of the intersection of general data trends and their relation to big data health topics is essential to the selection of source material appropriate to a systematic review.

Data engineering is being used in a public health capacity to configure, support, and maintain successfully next-generation health-related big data architectures. Engineering principles are being utilized to develop these systems in a fashion that balances a prototype and flexible access to many different data types within an equivalent framework while simultaneously ensuring high levels of data privacy, integrity, and availability. Some high-level characteristics of public health information systems that are designed to collect, store, and make available big data for public health research and practice have been put into play. However, community health and health disparities are dependent on geographically variable determinants that are hard to capture with standard health data measures. To accurately model and analyze the complex systems of behaviors, societal structures, and policy measures that give rise to these health disparities, diverse public health and community stakeholders require access to non-traditional and often highly-structured multi-sector data sources. Data engineering practices need to be widely integrated with emerging public health data frameworks so that these data are captured, maintained, and made accessible in standardized and high-quality ways. Such systems needed to securely scale so that data can be easily shared and interpreted, ensuring that analysis accurately reflects the geographic and community-based nature of the data. Data engineering practices must also actively promote and develop collaborative relationships between the engineers constructing the data frameworks, and the public health researchers and practitioners layering such data, to assure that the data captured will be accurate, informative, and useful through all stages of big data collection and analysis. It is concluded that community informatics for health disparities could benefit greatly by focusing on the key role of data engineering in the successful development and maintenance of a health surveillance system. At their core, successful public health interventions - whether enacted by government agencies, NGOs, or private enterprises - rely upon a lucid understanding of the health of the communities they serve, as well as the ability to effectively translate this understanding into informed, evidence-based policies and programs. Public health informatics - the management and analysis of public

health-related data - has long played a critical role in facilitating such translation efforts. However, recent advancements in data collection, storage, and analysis technologies have substantially expanded the range of public health-related data available for secondary analysis. This both complicates previous data management and analysis schemes, as public health informaticians must now contend with a broader set of big data formats, and presents new opportunities, as the development and deployment of new data analysis methods can potentially uncover new, actionable insights in the wider world of publicly available health data. It is within this rapidly changing environment that public health informaticians are turning to the data engineering community to optimize their toolkit of systems and strategies for handling such data, with an eye towards increasing both the usefulness and ease of use of data for a diverse set of public health stakeholders. At a fundamental level, actionable health-related insights depend on, and are communicable through, accessible, reliable, verifiable, and comprehensive representations of health data. As such, historically, "data practices" have long been appreciated as a crucial first step in any public health response. Inequities in the availability or quality of data profoundly impact the resulting public health outcomes; inaccurate or insufficient information can lead to incorrect understanding of disease causality or prevalence, hindering the appropriateness and timeliness of subsequent response. Public health data practices can impact the actions taken at all subsequent phases of the public health response, from the launch and targeting of initial supportive interventions, to the long-term progress of recovery activities. Substantiation of these assertions can be seen in a litany of disaster relief analyses, and additionally during the COVID-19 pandemic, as countries with more robust data infrastructures have been able to better manage and suppress case counts.

3.1. Role and Importance

[3] There is an increasing realization by Managers of public health programs that the use of a data engineering framework is critical for the successes and sustainability of prevention and community health surveillance system interventions. Modern community health surveillance systems are designed to be data savvy about what is happening in the communities that they function in. They are also aware that a large portion of the resources they deploy against a threat falls under this category. Data engineering frameworks address the processes of data collection, storage, and analysis leading to the development of health actionable outcomes. In essence it is the set of methods used to 'engineer' a sequence of data, the basis upon which it is turned into useful health data, historical data corresponding to a space-time window, which could be carefully curated, cleaned, geographically referenced then assembled in an analyzable form. In addition to these data, data engineering frameworks also address environmental, climatic and spatio-temporal sources of exogenous health data (although they may be analyzed and treated independently of this health data). Actors in a community health surveillance system are now, more than ever, aware of the fundamental importance of data saving in the strategy they develop to combat health emergencies. Among the growing range of possibilities are health professionals and epidemiological researchers, producing health forecasts; local communities, which data savvy signs of outbreaks, injury and fatalities; public authorities and policy-makers, developing resource allocation policies in the event of health deepening; media professionals and (therefore) the general public, who control the wave of social alarms; and finally the malicious agents who use data to organize health threats.

The importance of the prompt availability of this strategic intelligence when and where it is most needed, particularly in operational centers, drove the reliance of these practices toward the establishment of invested signaling links and data sharing protocols between the differing actors described. There are (at least) two ways in which good data engineering and sharing practices may enhance the epidemiological strategy proposed. It vastly expands the range of data available for the implementation of predictive analytics and disease modeling. Productive data engineering practice requires the development of

new technologies also (but not exclusively) requiring the investigation of solutions from within the realm of the community health surveillance system actors and necessitates a highly multicentric, interdisciplinary approach. Without sufficient attention being paid to these problems, even the most sophisticated and powerful algorithms and methodologies are likely to have a negligible impact on the efficiency and effectiveness of the public health strategies in place. On the other end, pursued in the way beneficial to the community as a whole, a strategic diagnosis of what data engineering solution within public health law will be useful is likely to be found as an incredibly powerful high-level action (an innovation complementary to those previously mentioned).
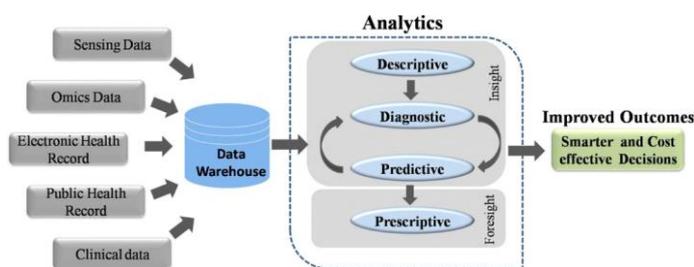

Figure 3. Big data in healthcare

3.2. Key Concepts and Techniques

The assessment focuses on time-series data processing and storage. This is a 2-part article. Part 1 enables healthcare managers and epidemiologists to comprehend the design and data requirements for a sparsely interconnected, 4-node health surveillance network. Part 2 matches the ESSENCE-State framework's detection and analysis functions to existing analytical data products. Part 2 includes four additional sections.

In the context of data engineering or analytics applied to public health, the text explores many data-related concepts or techniques in detail. Data cleaning is the essential but time-laboring process in data engineering. Data Integration is a broad process that contributes to data engineering, seeking effective ways to deal with different data and data source domains. Particularly, data fusion is thought of as a broad concept of data integration. Despite exceptional progress in big data research fields, better understanding in all aspects, frameworks, and data storage and processing are still challenging. Data visualization is important for health-related purposes because it can form health-related beliefs through health statistics interpretation. This article gives a comprehensive introduction of the application of data visualization in public health from the following aspects.

Data analytical technologies and models also progressed fast with the rise of data age and big data. Machine learning is a branch in data analytics that focuses on developing and implementing techniques and models that improve system performance. However, these analytical models are always complex and might consist of many state-of-the-art techniques. Data security and privacy become a significant health-related issue, nightmare public health threats. Security in the context of health surveillance is referred to as any method or technique used to safeguard data. On the other hand, data protection and privacy are related but separate security features. Data protection refers to ensuring data is not lost or corrupted, including data backup and recovery. On the other hand, privacy ensures that sensitive information is only seen by authorized agents or agencies. At present, developing solutions for the above two aspects are typically separate. This section discusses how data warehousing can be used to safeguard sensitive health surveillance data.

Despite these challenges, understanding the broader scope and applications of data engineering or analytics can enhance healthcare outcomes and research. The analytical or engineering features related to data are evolving fast and becoming more diversified and

multidisciplinary. Still, connecting their theoretical aspects with practical application is not an easy task. The central aim of the following is to provide a bridge between theoretical concepts and their practical applications in the health-related field. Or, the implementation of most of the public health framework is hard for public health professionals and other non-technical practitioners due to the hard terminology and methodology used. In this review, an effort is made to review, discuss and introduce those engineering or analytical steps that are related to data and/or required for the development of a robust data-driven health surveillance system, and introduce the broader aspects of data engineering and analytics from the health system. This will equip readers with a broader sense of how data engineering or analytics can benefit public health and/or health-related research.

Equation 2: Data Retention and Archiving

$$S_{archive}(t) = \sum_{k=1}^{p} S_k \cdot R_k(t)$$

where:

$S_{archive}(t)$ is the total storage requirement at time t,

S_k is the size of data from source k,

$R_k(t)$ is the rate at which data is stored from source k.

4. Frameworks for Data Engineering in Public Health

Public health informatics has strived to develop scientific methods to address the timely analysis, interpretation, and dissemination of information for enhancing the health of populations. This aims to compile a set of frameworks in data engineering in public health developed in the research community that are intended to allow enhanced data analysis by health managers and experts. These frameworks aim to provide new data analysis opportunities and insights from the data acquired from different sources, for instance medical consultations, environmental sensors, and search queries.

The frameworks focus on different aspects of the data acquisition, processing, and representation pipeline of the public health surveillance systems. The data engineering frameworks are categorized according to whether they address (A) data sanitization challenges, (B) data alignment requirements, (C) data fusion tasks, or (D) efficient data representational aids. The community public health surveillance systems vary greatly across different communities in terms of health challenges, needs, populous demographic, healthcare system, as well as collected data types. A challenging issue is to provide a framework that would be able to assist in devising a surveillance system in a way that would accommodate these differences and produce an optimal system. Scalability and flexibility to accommodate different types of data are important aspects to be taken under consideration when considering implementing a framework of a wide community application. It is important not only to provide a qualitative analysis of works but also perform a critical evaluation of their applicability. Therefore, each framework category is illustrated by an amalgamation of relevant works, followed by an in-depth analysis of the category, describing a few of the frameworks and pointing their strengths and weaknesses.

Figure 4. Conceptual framework for public health surveillance

4.1. Overview of Existing Frameworks

As one example, there already exists a range of developed public health data engineering frameworks. Although such frameworks indeed can benefit small to moderate size communities in capacities without resources for addressing their public health challenges, choosing or designing an appropriate framework could be challenging. Noticing the existing gaps in the matching of target communities with data engineering frameworks, a taxonomy for public health-related data engineering frameworks is presented. A broad range of existing frameworks of multiple roles and features was identified and categorized based on their functions, target communities, and data types to assist in navigating and selecting the right one for particular needs. Best practices applied in the implementations of existing work were also discussed to highlight elements that contribute to the success of frameworks, including usability, interoperability, and ways community input is incorporated. A unique contribution of this work is an integrative approach that couples a theoretical categorization with a practical discussion in community health surveillance and monitoring systems for practicing within data engineering frameworks. The presented taxonomy serves to facilitate broader and more effective uses of the growing diversity in public health data engineering frameworks and is generalizable for applications in other areas of public health informatics. Confidence in this framework classification is supported by a review of diverse case studies, demonstrating how different types of frameworks are matched with the unique challenges of different communities in order to effectively engineer their public health data. At the center of public health efforts on preparedness and response to health emergencies are disease surveillance and monitoring systems. Such systems require the real-time collection, analysis, and interpretation of health and health-related data to identify potential health threats. The framework similarly outlines these capabilities in the public health context. For timely discovery and intervention, such systems are enhanced by predictive modeling, such as how the framework is incorporated for spatiotemporal hot zone prediction in the context of infectious disease outbreak surveillance systems [4].

4.2. Comparison and Evaluation Criteria

This paper analyzed a plethora of frameworks and software regarding community health surveillance. In particular, this section focused on data engineering approaches, exploring tools and frameworks for health data collection, storage, and analysis. Using additional methods, journal articles and conference articles were systematically acquired for comprehensive review. The section presents disciplines and domains where it is likely to observe further research and applications. In addition, this section presents an in-depth comparison of the reviewed frameworks. Using concrete and context-specific quality, usability, and scalability criteria, an evaluation of the frameworks is provided on both qualitative and quantitative measures.

Comparison of the frameworks is performed regarding diverse criteria that are essential to ensure performance, usability, and adaptability of the surveillance system. Within- and cross-domain ability of the frameworks are examined based on different community health monitoring needs and case studies. Their system architectures are

analyzed to determine the trade-offs between ease of use and customizability. A user study case is presented to compare usability and user experience. With a quantitative evaluation approach, the system usability scale questionnaire is administered for a website and a mobile app. Substantiating qualitative findings, the quantitative results provide a more informed assessment of the impacts of the frameworks. Finally, in order to draw a comprehensive view, various other factors are discussed. A full understanding of the strengths and weaknesses of a framework is dependent on context-specific considerations. Therefore, not only does this paper sample systems from multiple disciplines, but also presents considerations on the possible application in various types of communities.
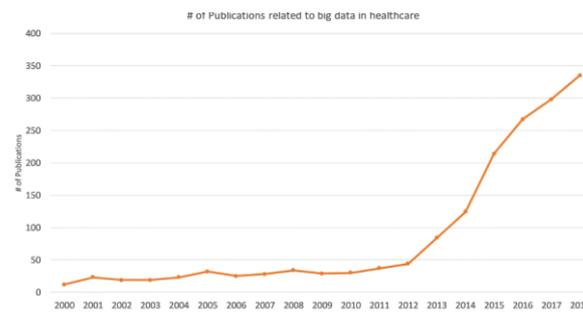

Figure 5. Big data in healthcare: management, analysis and future prospects

5. Case Studies and Applications

Case studies of the implementation of a new data engineering framework within Community Health Surveillance Systems are reviewed. Their setting, the challenges faced, explicit ways in which the application of new data engineering frameworks has been harnessed and examples of the resulting successful implementation of enhanced public health monitoring in each case are discussed. Diverse settings and challenges are highlighted (an urban slum, a rural population often comprising internal economic refugees and large transient populations in a rural district town). The need for stakeholder engagement and the differences between relatively "big tech" and local practitioners in the role of local academic collaborators are discussed. Measures of disease burden obtained and circumstances where the use of data engineering can demonstrate a clear return on investment are described.

At the Kilombero and Ulanga health and demographic surveillance system (HDSS) site, implemented in 1996 and expanded four times since to cover a population of more than 400,000, surveys are carried out, tri-tracked health facility and health centre information and image file-representation of blood slides are collected and village information relating to a further thousand people is maintained. This dataset provides unprecedented spatial granularity of malaria occurring within the site boundaries. One of the first challenges the site will need to address is the creation of a framework in HDSS which will enable broad access to public health information in a format usable by multiple internal and external organisations.
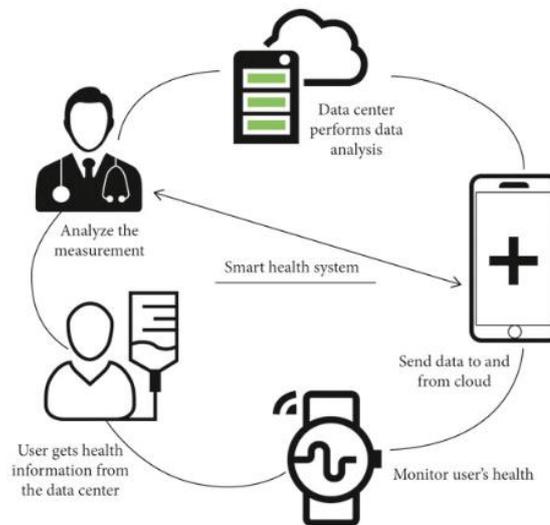
Figure 5. Applications in Healthcare Devices

5.1. Real-world Implementations

This section describes real-world implementations of data engineering frameworks in diverse community health surveillance settings. It provides a comprehensive view of several different examples, critically examining the specific objectives, methodologies, outcomes, best practices, challenges, and the significance of data integration and stakeholder collaboration in each. The intention is to illustrate the wide applicability of these frameworks in community surveillance contexts, and demonstrate the broader potential for collaborative data engineering in health surveillance . It is hoped that the best practice observation can provide useful guidance for new implementations, the discussion of limitations can provide insight into challenges that may be encountered, and the merged insights can illuminate the practical aspects of optimizing health surveillance systems.

On the one hand, it is a matter of public health practice that typical infectious disease surveillance systems are most effective when there is a good collaboration between different stakeholders. Nevertheless, different sets of data are collected by the health facilities and the community health centres, with the latter often collecting superior data largely because health facilities are frequently used by the wealthier section of the community. This situation is difficult to change in many developing world communities, but considerable improvements in mapping health could be achieved without any change in data collection through the shared use of these disparate datasets in a local health facility planning context. This situation calls for a mechanism for the data to be integrated, ideally in such a way that it is presented to all stakeholders in a uniform and comprehensible fashion.

5.2. Impact and Benefits

Community health can be improved by utilizing optimized public health approaches and principles of data engineering frameworks that operate community health surveillance systems. Some impacts and benefits are as follows: (i) greatly improved data accuracy throughout the surveillance system; (ii) much quicker time from the data collection to the decision of response activities; (iii) fully improved system capability to respond appropriately to any health threat; (iv) rapid addition of innovative techniques and inputs to health surveillance practices; (v) great economy of funding in comparison to the avoidance of health implications. Each impact and advantage are properly discussed and supported by various examples. Any community may easily improve its

health by applying relevant models with a very reasonable resource allocation and maintaining a multidisciplinary approach between various government stakeholders.

Implementation of a community health surveillance system model in the Temeke district of Dar es Salaam, Tanzania recently raised risks which were avoided by proper application of practices. A complex and thus improved health policy could have similar effects as the above neglect of health principles if there is no data to properly model health impacts. Necessary factors for the success of health policy and public health response results in the prevention and control of the assessed health risks. Such results can be easily achieved by implementing community health surveillance models and maintaining the community-based approach for data need identification.

Proper application of community health surveillance system practices may improve the health condition of any community. This widely recognized fact is reinforced by the consideration of the very resource-poor and conflict-ridden situation in Gulu district in the Northern regions of Uganda, where optimized application of health practices contributed to the shortening of the cholera epidemic for a few weeks and the total absence of malaria epidemics in 2003. Without any cholera outbreak it can be easily elevated under the described conditions of the malaria epidemic, yet the district managed to avoid it. Such cases also illustrate the increasing importance of taking into account the stochastic character and uncertainties about the possible impacts of health threats in conflict regions. The successful implementation of community health surveillance systems, as demonstrated in regions like Temeke district in Dar es Salaam, Tanzania, and Gulu district in Northern Uganda, highlights the profound impact such systems can have on public health. These systems not only enhance the accuracy and timeliness of data collection but also ensure rapid and informed responses to emerging health threats. In Gulu, for example, the optimized health practices contributed to the rapid containment of a cholera epidemic and the prevention of malaria outbreaks, despite the district's resource-poor and conflict-ridden conditions. This underscores the importance of a data-driven, community-based approach to health surveillance, particularly in areas with high uncertainty and risk. By focusing on timely interventions and proper risk identification, health outcomes can be significantly improved, even in challenging environments, showcasing the critical role of data engineering frameworks and collaborative governance in public health.

Equation 3: Resource Optimization for Health Surveillance

$$\sum_{k=1}^{p} a_{ik} x_k \geq b_i \quad \text{for each } i$$
$$x_k \geq 0$$

where:

C_k is the cost associated with resource k,

x_k is the allocation of resource k,

a_{ik} represents the amount of resource k required for task i,

b_i is the minimum requirement task i,

p is the total number of resources.

6. Future Directions and Emerging Trends

Community health surveillance involves continuous monitoring and evaluation of health determinants and outcomes to prevent serious illnesses, morbidity, and fatalities within a certain population. It has historic importance in epidemiology tracing back to the cholera epidemics of the 19th century and introduced the concept of data engineering for

the first time. Although vast improvements in medicinal sciences have been made since then, the basics of health surveillance, as in epidemic tracking and finding patterns in the data to prevent social infection or contamination, still bear the same premises. However, with the advent of digitized health data, prevention and precaution took on a new meaning. In the digital realm, the set of health data is massively broad and heterogeneous, from biorhythm plots to health check-up reports, medical diagnoses to genomics reports, all available at a mass storage system as long as it is connected on the internet.

In order to use this data effectively, sophisticated data engineering frameworks are crucial to transform, store, clean, harmonize, query, analyze, and visualize the data regularly. Due to the big data paradigm, health data grows in a way that far exceeds the regular formats of databases delay within organizations. With respect to certain determinants, it would be impossible to integrate every new technical advancement immediately; however, the framework must be designed with a high degree of adaptability in mind. Adhering to the fundamental premises, the aforementioned framework also addresses timely integration, storage, processing, and analysis of related data sources with regards to community health surveillance. Additionally, recommendations are provided for policy makers to further data integration within the healthcare system. In an era of vast improvements on enquiry prediction, personal data security is stressed more than ever. Thus, it should be noted that the adaptability of the framework would also consider ethical regulations. Moreover, a state of the art review is presented that, as far as this study is aware, addresses current and near future technological developments that would influence data engineering in relation to community health surveillance. Thus, this study aims not only to fulfill data engineers' need in the respective field, but also to prepare a road map for further better cooperation between; data engineer teams, each individual providing health organization, and public health professionals.

6.1. Technological Innovations

[5] Innovation in technology trends are shaping the future in the field of data engineering for public health practices. Advanced technological innovations, such as sensor technology, artificial intelligence, machine learning, internet of things (IoT) and big data analytics, provide real time data generation for health monitoring and readiness, and their efficient usage ensures management of generated data effectively and accurately . IT companies and computer engineers are making innovative breakthroughs in the medical and health sectors, and health-based mobile apps also aid and assist in monitoring vital records. In the last two calendar years, 232 health-related mobile apps have been innovated from IT companies. In addition, mobile and wireless health IT technology companies have already provided wireless health monitoring and management systems to hospitals and health clinics, monitoring real-time health status from anywhere in the world for admitted patients. The healthcare systems of the future will automate and optimize the more data-intensive tasks associated with healthcare, using epidemiological surveillance data to aid and guide public health officials . Health Analytics will support and develop a new science of health monitoring and rapid training and capacity building of the new class of "disease-analysts" that will be required to handle and manage the "thought-products" of this new system. The implications for public health policy and practice are endless and transformative. It will become possible to predict and pre-empt outbreaks. In recent years, advanced technological innovations in the field of community health surveillance have been successfully implemented in some cases to rapidly track and contain outbreaks. Public health systems protect and improve the health of individuals, families, communities, groups and population through health promotion, disease prevention, access to cures and rehabilitation. Globalization and urbanization lead to basic and traditional lifestyle transformations, and human morbidity and mortality rates significantly increase. As globalization and urbanization increase, health indicators

show significant declines in terms of lifestyle and infectious and communicable diseases. The healthcare systems of organizations will be adapted to the fast-growing health industry and become more services oriented with improved and enhanced services to monitor real-time health data. It will be the first community health surveillance system that will support and develop a database of health and medical examination records. Documentary Health and medical examination records that are especially available for electronic database systems in civic health centers can be monitored and analyzed. At the same time, Information, Education and Communication (IEC) activities and campaigns will also progressively be launched for officials and other practitioners in the community health system. A database of best practices will be set up to facilitate the communication, dissemination and meticulous adoption of new, opposite and successful community health surveillance experiences. In addition, human resources will equally constitute a pivotal resource input. Training and participatory aesthetic appraisal and therapy will focus on concretely improving the motivation, conceptual grasp and field skills of officials and other health practitioners in community health surveillance systems. Decentralized awareness and training programs in remote catchment areas will be likewise carried out while mobile delivery health posts will be developed to facilitate the organization and specialization of health care at the civic health center level and to efficiently coordinate a vast preponderance of settlers in the sites.

6.2. Research Opportunities

In maintaining the good health status of a community, it is important to implement a surveillance system. However, preventing and minimizing any kind of epidemic risks might bring various challenges in setting up a good surveillance system. One challenge to be addressed pertains to implementing new data engineering frameworks in the surveillance system by involving various parties including tech firms, data scientists, health organizations, and governments. Moreover, the adoption of the latest technologies will also be prioritized so that the health conditions of a population can be guaranteed well.

Apart from the challenge, there are also big opportunities to enhance the betterment of a surveillance system related to community health. First, there are numbers of research in which its findings can be deeply broadened such as emerging technology, big data analysis, Internet of Things (IoT), and drone application in healthcare. Therefore, it is worth it to elaborate these findings in order to solve complex emerging public health challenges. Second, the betterment of public health can be achieved by conducting the intersection of a variety of different research areas such as data engineering, demography, epidemiology, and computer networks. In addition, it is also important to focus on the contextual research impact by considering the health needs of diverse populations, as well as the local context of health such as health facilities, geographic topological conditions, and local health compliance. The method and experiment of a research can be extended by also conducting pilot studies, as well as experimental frameworks formulated in a way that is comparable through health outcomes or public health effects.

7. Conclusion

Introduction of information technology to health practices has liberalized a number of potentials in reengineering the diverse processes leading to systemic advancement. Following that trail, the health practices embraced significant transformations with the arrival of public health informatics to opt for informed decisions. Due to this, ongoing attempts by health professionals have strived to generate a health-surveillance framework with a focus on health data. However, the intricacies accompanied by this data need sophisticated data management. Thus, frameworks are now made within geeks showing the need for data engineering in line with the Bonferroni Framework.

Data engineering is surveyed considering the prevailing conceptualizations and the methodologies. There is discussion on information technology and its role in reengineering societal health practices. Health data is delineated in the instances of public health. Various endeavors have dealt with the designed framework for health data and herein-by providing an account of data engendering needs. Further endeavors are invoked for dynamic data engineering models that can adeptly transition amidst the data needs in plenum.

This essay presented an in-depth literature review on data engendering frameworks for a health-surveillance system. Scientometric quest is embarked, as a budding endeavor, is presented on the backdrop of enduring attempts to carve out a dynamic field with the substantive domain status. The conceptualizations of data engendering are manifold however open avenues to a budding review submerged subsequent to definitions. A systematic review of information technology applications in a health scheme pertaining to the societal course elicits the needs for data management within a health-surveillance system.

7.1. Summary of Key Findings

This research has explored the effectiveness of data engineering frameworks in optimizing community health surveillance systems. Frameworks are essential tools in addressing the existing challenges surrounding public health monitoring. Utilizing text analysis of a large set of open publications, essential themes in both public health and technology aspects of community health surveillance have been identified. Based on these themes, two data engineering frameworks have been proposed, one focusing on data sources and the other on data strategy. The results of a survey indicate that the vast majority of public health needs for effective monitoring can be met by utilizing the insights of the proposed frameworks. The open surveys have shown impacts on numerous health outcomes, including improved accuracy, timeliness, and completeness of data, as well as enhanced preparedness and decision-making. These results support the argument that data engineering plays a vital role in enabling various stakeholders to make better-informed decisions in effectively safeguarding public health.

Public health has traditionally been concerned with the timely analysis and monitoring of a broad spectrum of health determinants on community health. Community Health Surveillance Systems (CHSS) have continued this focus in digital form. Increasingly, the CHSS apparatus utilizes and produces large volumes of real-time heterogeneous data. Networks of public health organizations in several high-income countries have made recommendations for utilizing much of the data generated by mobile communication networks for real-time disease surveillance. Utilizing social media for public health surveillance is still a relatively new area. Similarly, there's indication of numerous requests for help from public health professionals for integrating difficult-to-obtain alternative data sources into analytical pipelines.

7.2. Future Trends

This subsection will examine future trends in (big) data engineering frameworks that aim to further optimize community health surveillance in the upcoming years. Technology is anticipated to further shift towards being recent and widely accessible, such as a seamless data stream in real-time from various health data sources, real-time data processing of e.g. machine learning algorithms, and real-time alarm generation. Methodology is also expected to change, such as an aggregator will be preferred between case-based health data and model-based health data, e.g. aggregation in the geographical and the temporal domain.

Currently data and surveillance technologies are manifold. Data is generated by online systems, mobile phones, social media websites, GPS devices, the internet of things, and sensors in high time resolution. Available unstructured health data is on the rise, such

as medical images and text. Surveillance systems are based on anytime-consultable low-cost big data infrastructure, distributed data mining, and machine learning approaches. Surveillance knowledge is distributed, relies on black-box models, and is often generated not only by health agencies, which bears the risk of biases. Best practices for first-response everyday workers interested in health surveillance to tackle an outbreak are still pending. Looking ahead on upcoming and therefore hopefully adjustable trends it is a comprehensive prerequisite for future surveillance systems. It investigates benefits, deficits, and necessary conditions for fair and ethical digital health data and health knowledge exploitation on a large scale. This data could be used in return for big data-driven modeling of illness spread, health intervention optimization, or fairness-aware data- and model-based personal health surveillance. Data engineering is likely to play a critical role in emerging practice.

References

[1] Vankayalapati, R. K., & Rao Nampalli, R. C. (2019). Explainable Analytics in Multi-Cloud Environments: A Framework for Transparent Decision-Making. Journal of Artificial Intelligence and Big Data, 1(1), 1228. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1228

[2] Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219959

[3] Chintale, P., Korada, L., Ranjan, P., & Malviya, R. K. (2019). Adopting Infrastructure as Code (IaC) for Efficient Financial Cloud Management. ISSN: 2096-3246, 51(04).

[4] Syed, S. (2019). Roadmap For Enterprise Information Management: Strategies And Approaches In 2019. International Journal Of Engineering And Computer Science, 8(12), 24907-24917.

[5] Mandala, V. (2019). Optimizing Fleet Performance: A Deep Learning Approach on AWS IoT and Kafka Streams for Predictive Maintenance of Heavy - Duty Engines. International Journal of Science and Research