

Big Data Applications in Vaccinology

Joseph E. Kasten, Pennsylvania State University, York, USA

ABSTRACT

The development of vaccines has been one of the most important medical and pharmacological breakthroughs in the history of the world. Besides saving untold lives, they have enabled the human race to live and thrive in conditions thought far too dangerous only a few centuries ago. In recent times, the development of the COVID-19 vaccine has captured the world's attention as the primary tool to defeat the current pandemic. The tools used to develop these vaccines have changed dramatically over time, with the use of big data technologies becoming standard in many instances. This study performs a structured literature review centered on the development, distribution, and evaluation of vaccines and the role played by big data tools such as data analytics, data mining, and machine learning. Through this review, the paper identifies where these technologies have made important contributions and in what areas further research is likely to be useful.

KEYWORDS

Big Data, Machine Learning, Pharmaceutical, Structured Literature Review, Vaccine, Vaccinology

INTRODUCTION

As this manuscript is being prepared, the world is struggling to produce a vaccine to prevent against the Covid-19 virus. Research on a wide variety of biological topics is currently underway and these build on countless other studies that have been completed over the preceding years. Many of these research efforts have relied on advanced analytical and data manipulation tools to both identify promising vaccine targets and to evaluate the efficacy of those already administered. The purpose of the present effort is to provide a structured review of the research literature describing the use of Big Data technologies in the development and evaluation of vaccines.

These analytical tools have been used to extract actionable information from vast databases of vaccine trial results (Ackerman, Barouch & Alter, 2017), develop an understanding of the unpredicted side effects of already developed vaccines (Iqbal et al, 2015), and predict the immunological attack points for future vaccine development (Conti & Karplus, 2019). In each of these scenarios, the use of Big Data technologies has helped to create vaccines more quickly, with more efficacy, and to identify additional uses for existing vaccines, thus increasing the pool of available vaccines without the expensive and time consuming process of developing them from the ground up.

To be clear, this is not a paper devoted to reviewing the technologies specifically utilized in the quest for a Covid-19 vaccine, though many of the tools described in this project will have undoubtedly been used in this effort. Rather, it is meant to describe the many Big Data-based tools that have been used in the recent past that have brought the field of vaccinology and immunology to the point where

DOI: 10.4018/IJBDAH.20210701.oa5

This article, published as an Open Access article on April 23, 2021 in the gold Open Access journal, International Journal of Big Data and Analytics in Healthcare (converted to gold Open Access January 1, 2021), is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

the development of a complex and critical vaccine such as those coming out for the Covid-19 effort can be developed, tested, and administered within the space of a single year. As in so many medical breakthroughs, what looks like a standalone discovery is actually the result of years of methodical, basic research that has paved the way for a specific advancement to happen in the manner that it has. Thus, this review of the literature describes the efforts that have led up to the present successes. The contribution made by a paper of this type is two-fold. First, it provides a framework within which researchers in the field, or those considering entering it, can place their research. This paper provides, for these researchers, a good overview of what is ongoing and, even more importantly, what still needs to be done. Secondly, for practitioners this paper represents a guide to those technologies that are being used or under development and might provide them with ideas about how to improve their processes or even develop new approaches to using big data in vaccinology.

BACKGROUND

When the non-epidemiologist thinks of vaccines, they think of flu shots and other immunizations given routinely to children and young adults. But, underlying these medical achievements is the need for processing and drawing understanding from huge datasets. Many of these Big Data processes are similar to those used in other industries, so from that perspective vaccine development and evaluation draws on tools that have matured in other areas such as finance and supply chain (Jordan, Dossou & Chang Jr., 2019). It is important, though, to understand how these tools are being employed, for as this understanding is more broadly dispersed throughout the data community, new applications can be more readily devised and more opportunities for protection from these illnesses and parasites can be developed.

Before proceeding with the study, it is important to define what is meant, in this effort, by Big Data technologies. This is important because of the wide range of understanding that exists about the definitions of Big Data and the tools and technologies associated with it. For the present project, a rather wide view of these terms is taken. This will allow the results to paint as clear a picture as possible of the state of the art and reduce the number of studies overlooked for terminological reasons. For the purposes of this study, the term “Big Data technologies” will include all forms of data analytics, machine learning, deep learning, and datamining. To reduce the complexity of the searches, the study does not use all possible terms describing these processes. However, the databases searched do a very good job of providing results that include other terms related to the basic terms listed above, so for example papers including terms such as “artificial intelligence” and “support vector machine” will be included even without those terms being used in the actual search expression. The terms used in this study are defined below:

- Data mining: using software and statistical models to search for patterns in large datasets.
- Machine Learning (ML): A broad category of algorithms that use data to train a model to make decisions. This includes a very wide collection of models and tools that are beyond the scope of this paper but can be reviewed in (Obermeyer & Emmanuel, 2016).
- Deep Learning: A branch of machine learning that enables tools such as Artificial Neural Networks (ANN). These can be trained to make decisions that support such processes as classification or speech recognition. These are closely related to big data technologies because they require very large datasets to train and verify the algorithms.
- Data Analytics: A broad term covering all of the statistical, visual, and mathematical tools and techniques commonly used to analyze and draw meaningful information from big data repositories.

While this is not an exhaustive list of technologies related to Big Data, it provides a good starting point with which to examine the literature surrounding the use of Big Data technologies in

the development and evaluation of vaccines. The following section details the methodology used in this structured literature review.

METHODOLOGY

As this is a structured review of the literature surrounding the use of Big Data technologies in the vaccine development and evaluation space, it is crucial to employ a methodology that ensures rigor and provides results that can be replicated. The present study follows the process outlined by Briner & Denyer (2012). These authors provide a five-step process to follow when completing a study of this sort:

- Identify the research question(s).
- Locate and select relevant studies
- Critically appraise the studies
- Analyze and synthesize the findings
- Disseminate the findings

The study seeks answers to the following research questions:

- For which vaccine development/evaluative processes are big data technologies (as defined in this study) being used to improve or analyze?
- For how long have big data technologies been used to improve or analyze vaccine development/evaluative processes?

The first research question captures the overarching reason for the study. The second research question, once resolved, will provide the reader with an understanding of the maturity level of the Big Data/vaccine combination. For researchers, this provides some insight into the availability of new research directions, while practitioners in the data science, pharmaceutical, and regulatory fields will be able to evaluate how deeply embedded these tools are, and that might lead to a better understanding of how they could fit into the overall vaccine development process.

In order to facilitate the structured review, certain “ground rules” must be in place to both ensure repeatability and to limit the scope of the review. Therefore, the following stipulations apply to this study:

- The literature included in the study will consist only of published works such as journal articles, conference proceedings, book chapters, etc. White papers, works-in-progress, and other so-called “gray” literature, while quite possibly making important contributions to the field are excluded from the study. These might be an interesting area of inquiry for future study.
- Literature generally aimed at the practitioner, rather than researchers, is also excluded from the study. As above, there is likely useful information in these publications, but the point of the current effort is to locate the areas currently being researched and developed in the academic literature. Again, this would be a fruitful area for future research.

Searches were carried out in many of the major electronic databases that provide coverage for the vast majority of scholarly journals in the fields of data analytics, information technology, and healthcare. The databases analyzed were ABI/Informs, Emerald, IEEE Explore, JSTOR, Science Direct, Scopus, Springer, Taylor & Francis, Web of Science, Google Scholar, PubMed, CINAHL, and ACM. The searches were conducted by searching the metadata (title, abstract, keywords) using the following search terms in all combinations: big data, data analytics, deep learning, machine learning,

data mining (datamining), vaccine, and immunology. These search terms cover a great deal of the topic area, but this study does not claim to be exhaustive. However, as the searching continued and the amount of duplication among the various databases increased, it seems that this search protocol identified a large percentage of the available literature. Figure 1 displays the study protocol.

The returns from the initial set of searches numbered in the thousands. The next step in the process was to identify which of the articles in the search results fit the needs of the study. In some cases, the article title provided sufficient evidence that the article fit the study. However, for many articles the title was sufficiently vague that the abstract, the introduction, and in some cases the findings and conclusion, were analyzed to finally determine that the article was indeed about the use of some form of Big Data technology in the field of vaccinology. After culling the studies that do not focus their attention on the use of Big Data technologies on the development and application of vaccines, there remained 325 papers. The composition of these papers and the research themes contained therein are described in the Findings section.

FINDINGS

There are two subsections in the Findings section. The first subsection provides descriptive statistics for the dataset collected using the protocols described above. The second subsection provides an analysis of the various themes and subthemes that are evident in the academic literature.

Descriptive Statistics

Of the 325 documents included in this study, there are only three types of documents represented: journal articles, conference proceedings, and edited book chapters. No theses, dissertations, or other types of documents were uncovered. Table 1 displays the breakdown of documents by document type.

The proportion of journal articles to conference proceedings and book chapters varied by theme but in each case the number of journal articles far exceeds the combination of proceedings and book chapters. In only two cases, the group of documents focused on pharmacovigilance and reverse vaccine design, were there no conference proceedings or book chapters. These trends will be discussed further in the Discussion section.

There was no temporal restriction put on the search terms in order to gain an understanding of when the association between Big Data technologies and vaccinology first began and how quickly it has grown. Figure 2 provides a graphical display of the growth of Big Data technologies and vaccine-related literature. A few early papers, the first being published in 1998, made some initial connections between the two concepts, but it wasn't until approximately 2008 that the body of literature began

Figure 1. Structured literature review protocol

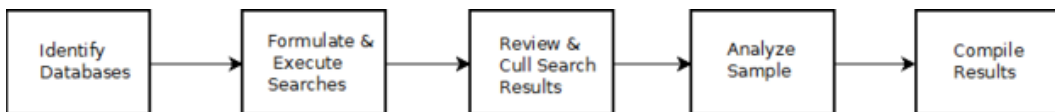
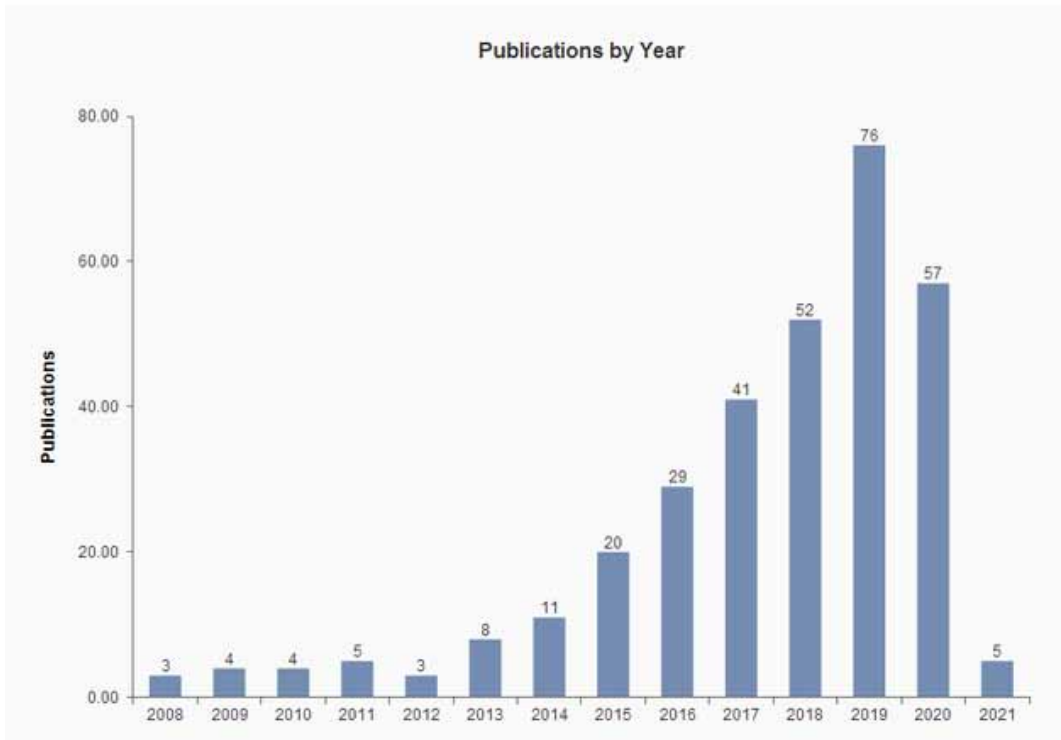


Table 1. Documents found by type

Document Type	Number Found
Journal articles	282
Conference Proceedings	34
Book Chapters	9

Figure 2. Number of publications by year



to build in earnest. A sharp increase in publications begins in 2013 and peaks in 2019. The searches were all complete by the end of November, 2020, so it is likely that the final values for 2020 will be somewhat higher. Also, a few publications scheduled for 2021 are included in the data, giving a preview of the upcoming activity.

Research Themes

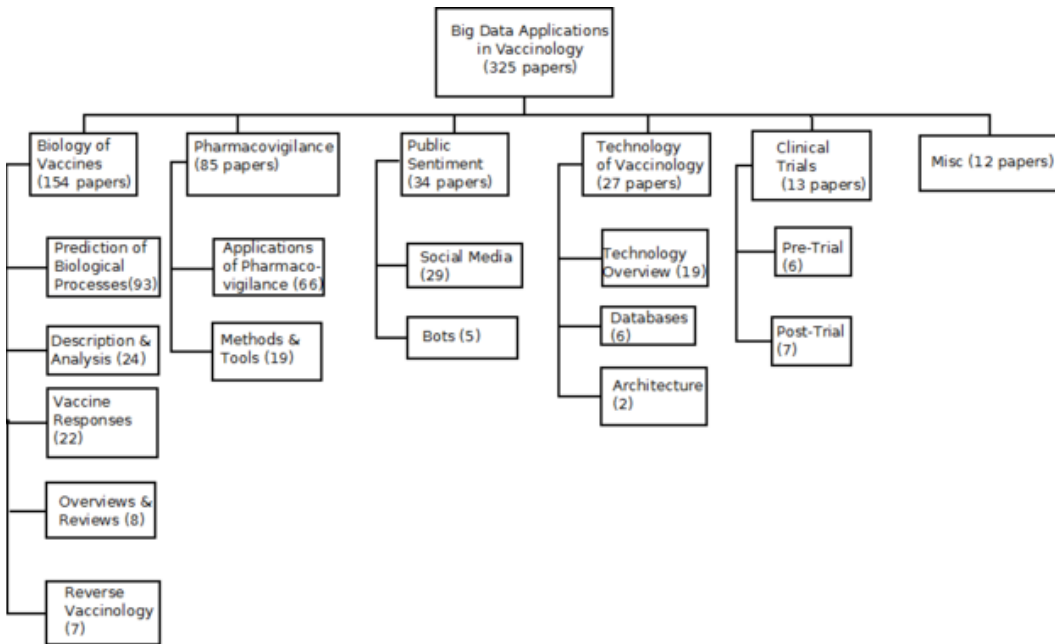
During the analysis of this collection of literature, a number of distinct research themes emerged. However, it also became apparent that even within these rather well-defined themes, there was a need to provide additional clarity regarding their individual emphases. To fulfill that need, many of the themes described in this paper also contain a number of sub-themes, or subdivisions, within them. These will provide a greater degree of specificity in providing an explanation of the six research themes identified. A graphical depiction of these research themes is presented in Figure 3.

As the scale of the current study is rather large at 325 items, it is impractical to describe each of these papers in any detail. However, it is important that some level of detailed explanation is provided to convey the importance of this body of literature. As a compromise, a selection of works from each area of concentration (sub-themes) is highlighted and discussed and a larger selection is referenced for each sub-theme in the tables that follow each sub-section.

Theme: Biology of Vaccines

The first and largest theme in the body of literature centers on the actual biology of vaccines and the antigens they protect against (152 papers). Within this theme can be found a number of sub-

Figure 3. Research themes and sub-themes



themes, the largest of which deals with the prediction of various biological interactions among the various components of antigens and antibodies. Within this sub-theme, the largest thread centers on the prediction of how certain biological entities will interact. Epitopes, sometimes known as antigenic determinants, are the specific place on the antigen that an antibody can attach to in an effort to neutralize the attacker (Khanna & Rana, 2017). Kar et al (2018) describe various methods to identify likely immunogenic amino acid sequences such as artificial neural networks (ANN) and support vector machines (SVM). Rosyda, Adjil & Setiawan (2016) use a cascading neural network to predict the epitope region on the P24 protein on the HIV virus. A deep neural network tool is used by Sher, Zhi & Zhang (2017) on a number of standard protein datasets and was shown to achieve substantially better performance than other accepted prediction tools. Additionally, Solihah, Azhari & Musdholifah (2020) couple an oversampling technique with a SVM to develop a conformational epitope prediction model and a method called BPairwise is developed by Zhang & Niu (2010) to predict variable-length epitopes.

Epitopes are composed of proteins, and the next sub-theme focuses on the identification and prediction of virulent proteins. Munteanu et al (2014) create a new model using linear discriminant analysis to discriminate the 3D structure of lectins (carbohydrate-binding proteins) from a collection of protein structures with unknown functions. These are important in developing therapies and vaccines for a wide range of cancers, parasitic infections, and other diseases. An ensemble of SVM are used by Nanni & Lumini (2009) to predict the functions of virulent mechanisms in pathogens by extracting the features directly from the protein's DNA. Singh, Singh, and Sisodia (2019) approach the limiting factor in the utilization of ML in protein prediction, the lack of adequate training datasets, by developing a composite model for a multitask learning framework. This will facilitate the building and verification of training datasets, thus opening additional pathways for protein identification and cataloging. Lastly, El-Manzalawy and colleagues (2016) tackle the problem of predicting the existence and character of surface proteins on the malaria parasite using a semisupervised ML algorithm.

Peptides are composed of a small number of amino acids chained together and are distinguished from proteins by consisting of a smaller number of amino acids, though the cutoff between the two is somewhat arbitrary. Peptides bind to the major histocompatibility complex (MHC), which are cell surface proteins that bind peptides to the cell surface so they may be recognized by T-cells (Janeway et al, 2001). They are important in the identification of cells that should be considered part of the organism or possible antigens that need to be attacked. Vang and Xie (2017) use ANN technology to predict the binding tendencies of these peptides. Boehm et al (2019) use a similar approach, but use public databases describing MHC immunopeptides to train random forest classifiers, a specialized form of ML. Liu et al (2020) use a similar approach to identify these MHC binding peptides in the course of vaccine development for SARS-CoV-2.

Predicting antigen/antibody interaction is the next area of interest in this subtheme. This is the next step after the surveys of epitope and peptide behaviors described above. These studies are also applicable to vaccine development beyond those of infectious diseases, such as those for certain cancers. For example, Smith et al (2019) use ML techniques to analyze the behavior of tumor antigen immunogenicity. Conti and Karplus (2019) use ML to uncover a universal HIV vaccine that would neutralize multiple variants of the disease. Focusing on specific antibodies, IgA and IgG, Khanna and Rana (2017) use an ensemble ML approach to predict the existence and amount of these in a sample dataset. Successful prediction of these antibodies could help formulate treatments and vaccines.

The final thread within the prediction sub-theme contains studies that seek to predict the performance, or sometimes lack of performance, of vaccines once they are administered. Examples of these include Parvandeheh et al (2019) study using ML to predict the antibody response to the influenza vaccine, a study by Lee et al (2015) that seeks to understand exactly how the H1N1 influenza A virus provides protection for its human recipients, and the effort by Hemedan et al (2020), which attempts to predict the existence of vaccine-derived poliovirus outbreaks.

The second sub-theme in the biology theme is the use of Big Data technologies in the description and analysis of virus types and components. An important example of this is the use of a novel data-mining algorithm to identify the sexual development cycle of the malarial parasite in order to facilitate improved vaccine design. Mining of huge libraries of Kröhnke pyridines to find keys to the production of an arenavirus vaccine is another important application of these tools. Pyridines are organic compounds that are important in the creation of pharmaceutical compounds. Other efforts in this sub-theme involve the use of datamining samples taken from HIV-positive people to understand the exact immunological responses of their individual immune systems to see how they tried to fight off the virus, even if unsuccessfully, and a similar effort to mine the compounds from African flora with potential to inhibit the activity of the Ebola virus.

The analysis of the actual response to a vaccine, rather than a prediction, is the content of the third sub-theme. In some cases, the actual mechanism by which vaccines provide protection is not completely understood, so post-vaccination analysis can be useful in subsequent vaccine development as well as defining adjuvant therapies. Flanagan et al (2013) derive their understanding of a vaccine response by profiling the reaction of the entire immune system rather than only the cellular level of response. Pittala, Morrison, and Ackerman (2019) perform this process on humoral immune responses to HIV. Chaudhury et al (2020) analyze the effects of adjuvant formulation on vaccine-induced immunity.

Sub-theme four includes those authors whose goal is to provide an overview of the field, either as an introductory work or to provide a footing for future research. Some examples of these works include Cotugno et al (2015) effort to review techniques for gene expression analysis as part of the search for a functional cure for pediatric HIV, Izak, Klim and Kaczanowski's (2018) review of the use of bioinformatics in the definition of host-parasite interactions in the fight against malaria, and Olafsdottir, Lindqvist, and Harandi's (2015) work to define the tools used in the derivation of molecular signatures of vaccine adjuvants. These papers and others like them provide a suitable underpinning of the use of analytic tools as applied to very specific problems, just as the current project provides a larger portrait of the field for a wider view of the use of these technologies in many aspects of vaccinology.

The last sub-theme in this section is labeled “reverse vaccinology,” which is a term that has gained some traction over the past few years. In reality, much of the material reviewed in the current theme falls under the heading of reverse vaccinology (RV), which simply means that the search for epitopes and their compatible antigens occurs *in silico* (on computers) rather than *in vivo* (in living organisms). The reason for setting these works aside into a separate sub-theme is because their language represents a specific term within the field of vaccinology. This terminology might be used to a) define a specific project or effort or b) be used as a specific search term by subsequent researchers. Therefore, the topics covered in this sub-theme include the evaluation of different classification algorithms to predict protein sequences (Heinson et al, 2019), a comparison of open source RV tools for bacterial vaccine antigen discovery (Dalsass et al, 2019), and a description of how ML and RV were used in the COVID-19 vaccine development process (Ong et al, 2020). Table 2 presents the works in the Biology of Vaccines theme.

Theme: Pharmacovigilance

Pharmacovigilance is the term given to those activities which allow the government and the vaccine manufacturer to monitor the performance and side effects of a vaccine, or any drug, after it is approved and released. In many cases, the monitoring is done with data submitted to the Vaccine Adverse Event Reporting System (VAERS) which is co-administered by the Centers for Disease Control (CDC) and the US Food and Drug Administration (FDA). VAERS reports can be submitted by anyone, but healthcare providers and drug manufacturers are required by law to submit any adverse events they become aware of. This data set is publicly available (vaers.hhs.gov) for viewing or analysis.

With this as background, there are two sub-themes evident in the pharmacovigilance theme, applications of Big Data tools to pharmacovigilance and the methods used in performing these tasks. The applications of Big Data to this important process form a rather large group of papers (65). A

Table 2. Papers Included in the biology of vaccines theme

Sub-Theme	Author(s)
Prediction of Biological Processes	<p>Epitopes: Kar et al, 2018; Khanna & Rana, 2020; Kim et al, 2019; Sher, Zhi & Zhang, 2017; Solihah, Azhari & Musdholifah, 2020; Rosyda, Adji & Setiawan, 2016; Zhang & Niu, 2010</p> <p>Proteins: El-Manzalawy et al, 2016; Munteanu et al, 2014; Nanni & Lumini, 2009; Singh, Singh & Sisodia, 2019</p> <p>Peptides: Boehm et al, 2019; Degoot, Chirove & Ndifon, 2018; Henneges, Huster & Zell, 2009; Liu et al 2020; Vang & Xie, 2017</p> <p>Antigen/Anitbody: Conti & Karplus, 2019; Khanna & Rana, 2017; Kim et al, 2018; Nagpal et al, 2018; Smith et al, 2019</p> <p>Vaccine Effectiveness: Cotugno et al, 2020; Hemedan et al, 2020; Lee et al, 2015; Parvendeh et al, 2019</p>
Description & Analysis	Kwofie et al, 2019; Miranda et al, 2018; Tumiotto et al, 2017; Young et al, 2005
Vaccine Responses	Chaudhury et al, 2020; Flanagan, Noho-Konteh & Dickinson, 2013; Pittala, Morrison & Ackerman, 2019; Tomic et al, 2019
Overviews & Reviews	Cotugno et al, 2015; Izak, Kim & Kaczanowski, 2018; Olafsdottir, Lindqvist & Harandi, 2015
Reverse Vaccinology	Dalsass et al, 2019; Heinson et al, 2019; Ong et al, 2020

sample of these studies is characterized by studies that determine the safety of a particular vaccine (Myers et al 2017; Haber et al, 2019; Haber et al, 2020). Other topics of study within this sub-theme include studies that attempt to develop a profile of those who might be at risk with a particular immunization, such as a particular pneumococcal vaccine (Yildirim et al 2020; Iqbal et al 2015), and studies that seek to link vaccine benefits with ambient conditions such as air pollution (Liu et al, 2020).

The second sub-theme discusses various methods used to perform the pharmacovigilance process. Many of these authors describe improved methods of sharing data, such as Briggs' (2017) study on the Standard for Exchange of Nonclinical Data (SEND). Others support the use of ontologies as a means to improve the usage of data across disciplines and nations (He, 2014). There are also authors who describe various data sources that can aid in pharmacovigilance such as the Influenza Research Database (Zhang et al, 2017; Trifiro, Sultana & Bate, 2018). Lastly, a number of researchers provide methodological improvements for mining these databases such as the effects of timing (Berendsen et al, 2016; Svanström, Callrés & Hviid, 2010; Cai et al, 2017), sample stratification (Woo et al, 2008), and subgroup analysis (de Bie et al, 2012). Table 3 summarizes the papers in the Pharmacovigilance theme.

Theme: Public Sentiment About Vaccines

The only way in which vaccines can be effective is to get them into the arms of the people they are meant to protect. In some cases, these immunizations are required by law, as in the case of children in the US who must be inoculated against a host of childhood diseases before attending school (though there exist many exemptions from this requirement). However, many vaccines are administered only when the patient consents. In some cases, public sentiment against the use of vaccines can thwart the effectiveness of vaccines because of a low acceptance rate. Whether the lack of trust is based on fear, misinformation, or other factors, the ability of a pharmaceutical firm, and in some cases the government, to engender trust in the public is the key to a successful inoculation program. This theme includes those studies that seek to understand the role of public sentiment in vaccine usage and how this public sentiment is influenced.

This theme consists of two sub-themes, which are closely related. The first sub-theme concerns the role of social media in the shaping and movement of public sentiment both for and against vaccines. Certainly, the largest number of these studies focuses on the use of the Twitter platform as a communication and influencing tool. Studies that seek to understand the role of tweets in the decision to take a specific vaccine make up the largest portion of this group, with studies on the usage of influenza (Krittanawong et al, 2017; Kagashe, Yan & Suheryani, 2017; Chan, Jamieson & Albarricin, 2020) human papillomavirus (Du et al, 2017; Massey et al, 2016; Dunn et al, 2017), and the measles-mumps-rubella vaccine (Pananos et al, 2017) leading the field. Other studies in this sub-theme center on the relationship between Twitter and the overarching public opinion on vaccinations (Tavoschi et al, 2020), partisan politics' impact on vaccinations (Walter, Ophir & Jamieson, 2020), vaccines during pregnancy (Martin et al, 2020), and vaccination-related autism (Tomeny, Vargo & El-Toukhy, 2017).

Table 3. Papers included in the pharmacovigilance theme

Sub-Theme	Author(s)
Applications of Big Data to Pharmacovigilance	Haber et al, 2020; Haber et al, 2019; Iqbal et al, 2015; Liu et al, 2020; Myers et al, 2017; Yildirim et al, 2020
Methodological	Berendsen et al, 2016; Briggs, 2017; Cai et al, 2017; de Bie et al, 2012; He, 2014; Trifiro, Sultana & Bate, 2018; Svanström, Callrés & Hviid, 2010; Woo et al, 2008; Zhang et al, 2017

Not all studies center on Twitter. The content generated on other websites is also of concern. Getman et al (2018) study the effect of information from other sources such as Wikipedia and the New York Times on vaccine hesitancy, Japanese researchers study the impact of large Japanese datasets on vaccine hesitancy (Nawa et al, 2016; Okuhara et al, 2018; Okuhara et al, 2019a) and Meyer et al (2019) used the Canadian Broadcasting Corp. website as a target. Some researchers also look at the impact of newspapers on the vaccination decision (Okuhara et al, 2019b). Other researchers examine social media as a tool to promote on-time vaccinations for children (Chandir et al, 2018; Bell et al, 2019).

The second sub-theme takes as its target the use of bots in the creation of social media posts to influence the vaccination decision. Yuan, Schuchard & Crooks (2019) study the communication patterns of pro- and anti-vaccine users and the role of bots in shaping those patterns. Kudugunta & Ferrara (2018) study the problem of detecting bots so that they might be identified and their content flagged or removed from the platform. Table 4 contains the papers contained in this theme.

Theme: Technology of Vaccinology

The specific technologies involved in vaccinology are the subject of this theme. The first sub-theme covers the various overviews and reviews that are prevalent in these papers. These reviews vary in specificity from taking a very wide view to something very narrow, but in all cases they represent a very useful resource for gaining an understanding of the role of information technology, and Big Data tools specifically, in the development of vaccines. de Sousa, de Menezes Neto, and de Brito (2013) explain the challenges of mining the data generated by the DNA sequencing processes used to predict protein behavior. Oberg and colleagues (2015) present some of the lessons learned in dealing with high-dimensional, high-throughput data as applied to vaccinomics. Rodrigo and Luciana (2019) discuss the applications of next generation sequencing (NGS) in studying host-pathogen interactions. Finally, and aptly, Vaishya et al (2020) discuss the role of artificial intelligence in the fight against Covid-19.

Databases play a huge role in computer-aided vaccinology. Because the huge amounts of data generated by tools like NGS need to be formatted in a specific way to be ready for various types of analyses, the database technologies utilized can make a significant impact on system performance (Davies et al, 2007). Tomic et al describe the FluPRINT dataset to enable large studies that explore the underlying concepts of the influenza virus from a cellular and molecular level. To bridge the gap between the immunology and ML communities, the Dana-Farber Cancer Institute has developed a tool to provide standardized datasets to enhance the compatibility of ML-based research in virology and immunology. Finally, the National Institutes of Health (NIH) has developed the NIH Immune Epitope Database, which contains curated datasets created to enhance the prediction of peptide/MHC binding.

Finally, the third sub-theme includes papers that describe certain architectural aspects of systems built specifically for reverse vaccinology research. For example, Dharayani and colleagues (2019) describe a MapReduce-based architecture to run the BLAST Algorithm, an open source tool for

Table 4. Papers included in the public sentiment theme

Sub-Theme	Author(s)
Social Media	Chan, Jamieson & Albarracin, 2020; Du et al, 2017; Dunn et al, 2017; Getman et al, 2018; Kagashe et al, 2017; Krittanawong et al, 2017; Massey et al, 2016; Martin et al, 2020; Meyer et al, 2019; Nawa et al, 2016; Okuhara et al, 2019a,b; Okuhara et al, 2018; Pananos et al, 2017; Tavoschi et al, 2020; Tomeny, Vargo & El-Toukhy, 2017; Walter, Ophir & Jamieson, 2020
Bots	Bell et al, 2019; Chandir et al, 2018; Kudugunta & Ferrara, 2018; Yuan, Schuchard & Crooks, 2019

comparing biological sequence information. Table 5 displays some of the papers included in the Technology of Vaccinology theme.

Theme: Clinical Trials

As the world has witnessed over the past year, clinical trials are at the heart of vaccine evaluation and approval. Without data from a sizable group of volunteers who have received the vaccine, there would be no basis for the approval and distribution of a vaccine, regardless of how well it performs in the lab or in a computational simulation. In this theme, the literature describes the use of Big Data technologies in both the pre-trial and post-trial phases of the process. In terms of pre-trial uses, Big Data tools have found their way into the process of matching patients to study acceptance criteria (Henderson, 2016), supporting sequential patient recruitment, sometimes a result of paused trials (Zame et al, 2020), and using deep learning tools to simplify the process of understanding eligibility criteria (Chuan, 2018). Moreover, Morgan (2019) uses analytical tools to pinpoint, and reduce, the costs of clinical trials, which are more expensive to run than most people expect.

During and after the clinical trial is complete, there are many opportunities to utilize Big Data technologies. Lopalco and DeStefano (2015) use datamining techniques to evaluate both Phase 3 trial findings and post-licensure pharmacovigilance to evaluate the safety and efficacy of a vaccine. Machine learning tools are useful when evaluating the actual protection value of the chicken pox and shingles vaccines (Gilbert & Luedtke, 2018; Ackerman, Barouch & Alter, 2017). Large database access and mining tools make analysis of post-trial data easier and more transparent, as well as making the results of these analyses more useful (Zung, 2019). The papers included in the Clinical Trials theme are displayed in Table 6.

Theme: Miscellaneous

This theme consists of Big Data tools that are used beyond those activities outlined earlier. These applications, though not numerous enough to be given the status of theme, nevertheless are an important part of the vaccine process and should be enumerated. Within this theme, there are a number of interesting areas of research. Agarwal et al (2020) demonstrate the usefulness of deep learning processes to develop a profit function for the pharmaceutical firm. The manufacturing of vaccines is the focus of a number of efforts. Wu et al (2019) use a ML approach to perform pose estimation to

Table 5. Papers included in the technology of vaccinology theme

Sub-Theme	Author(s)
Overview	Banerjee, Basu & Nasipuri, 2015; Bragazzi et al, 2018; de Sousa, De Menzes Neto & De Brito, 2013; Lalmuanawma, Hussain & Chhakchuak; Oberg et al, 2015; Rodrigo & Luciani, 2019; Vaishya et al, 2020
Databases	Altmann, 2018; Tomic et al, 2019; Zhang et al, 2011
Architecture	Dharayani et al, 2019

Table 6. Papers included in the clinical trials theme

Sub-Theme	Author(s)
Pre-Trial	Chuan, 2018; Henderson, 2016; Morgan, 2019; Zame et al, 2020
Post-Trial	Ackerman, Barouch & Alter, 2017; Gilbert & Luedtke, 2018; Lopalco & DeStefano, 2015; Zung, 2019

support fully automatic mosquito salivary gland extraction, a step necessary to creation of malaria vaccine. Neural network tools are used by Yu et al (2019) to determine the fertility of chicken embryos, important for the manufacture of a variety of vaccines.

The sophisticated supply chain necessary to maintain the proper conditions while vaccines are distributed, usually focused on the temperature of the vaccine, benefits from the use of Big Data tools. Yong et al (2020) suggest a blockchain-based approach to vaccine supply chain management using ML tools to analyze the chain’s performance. Sujaree (2019) suggests the use of the max-min ant system (MMAS) algorithm to design a vaccine cold chain. Bhattacharjee et al (2015) develop a system to integrate the various sources of data generated during the transportation of vaccines in an effort to provide a more comprehensive portrait of the transportation process.

The topic of immunization protocols is addressed by researchers using Big Data tools. Chen et al (2020) use ML to develop clinical decision support system for shingles vaccination and Hovav et al (2017) take a similar approach using health care analytics to support an influenza vaccination program. Finally, Bhatti et al (2018) utilize datamining to help create a recommender system to support optimal coverage of vaccinations and Clark et al (2011) propose the use of patent data mining to find vaccines that can be repurposed for other uses to avoid the time and expense of new vaccine development. Table 7 presents the papers categorized in the Miscellaneous theme.

DISCUSSION

The results of this study illustrate the wide variety of uses to which Big Data tools have been put in the development and deployment of vaccines. These tools have been used to quicken the development of vaccines as well as making them more effective, safer, and less costly. Within these results, there are a number of interesting trends and issues that necessitate further discussion. It is apparent by the preponderance of the literature that the use of these tools within the biological activities is well accepted and a very appropriate application. These tools are designed for the purposes of analyzing combinations of many variables such as the combination of epitopes and antibodies. Likewise, the second largest piece of the literature centers on the use of datamining technologies to evaluate the huge databases that contain data on adverse outcomes from vaccine use. With the other themes being much smaller, it seems that the research world has either not seen the need for these tools in these various areas or that they have not yet addressed them.

As a tool to understand and plan Big Data implementations, there have been many “stacks” put forth by authors across the Big Data and analytics landscape. These stacks are frameworks developed to provide a convenient roadmap for researchers, students, and practitioners to follow when designing, analyzing, or learning about Big Data implementations and how they can be deployed. For example, Frampton (2018) describes a Spark-based stack for the creation of an open source Big Data stack, Lu et al (2018) present a Deep Learning Over Big Data (DLoBD) stack that provides a more specific approach to combining Big Data with high performance computing and compare performance across multiple configurations, and Sakr (2017) analyzes the performance variations between Apache and Hadoop – based stacks. For the present study, a Big Data stack might be a valuable lens through which

Table 7. Papers Included in the miscellaneous theme

	Author(s)
Miscellaneous Theme	Agarwal et al, 2020; Bhattacharjee et al, 2015; Bhatti et al, 2018; Chen et al, 2020; Clark et al, 2011; Hovav et al, 2017; Montagna et al, 2020; Pennisi, Russo & Pappalardo, 2018; Sujaree, 2019; Wu et al, 2019; Yong et al, 2020; Yu et al, 2019

to view the contributions of some of the authors. While it is infeasible to evaluate each contribution through this lens, a few examples of each type will be useful. The simplified stack in use here consists of four layers:

1. **Data Storage:** technologies that store huge masses of data such as Hadoop or Amazon S3.
2. **Integration & Ingestion:** technologies that perform the ingestion and management of data from their sources (Ex. Stitch, Apache Kafka) as well as provide Extract, Transform, & Loading (ETL) services (Ex. Python).
3. **Processing:** technologies to perform the calculations necessary to address the research questions (Ex. Apache Spark, Map/Reduce, Tensorflow).
4. **Analytics:** technologies to perform analytical calculations as well as provide visual analytics and dashboards (Ex. Tableau, R).

With this framework in place, Table 8 provides some examples of tools in these layers used to improve vaccinology. It should be noted that it is unrealistic to describe each study as only using one of these layers of technology. In reality, each portion of the stack is being used in almost every effort. For example, when an author describes the use of a machine learning tool to predict the binding propensity of a peptide, it is assumed that they are also using some form of data storage tool, probably employing ETL or other pipelining method to improve the flow, and analyzing results on a

Table 8. Technology application with reference to big data stack

Layer	Author	Tool	Results
1. Data Storage	Zhang et al, 2016 Miranda et al 2018 Massey et al, 2016	Influenza Research Database Kröhnke Pyridine Library multiple data collection tools	Describe use of DB to enhance research collaboration Demonstrate value of DB in research of arenavirus Analyze Twitter data to quantify HPV communications
2. Integration & Ingestion	Briggs, 2016	Standard Exchange for Nonclinical Data	Describe use of common data format to reduce the need for certain ETL activities
3. Processing	Dharayani et al, 2019 Flanagan et al, 2013 Degoot, Chirove & Ndifon, 2018	BLAST algorithm on Map/Reduce platform datamining tools machine learning	Decreased time required to search for genetic anomalies Improve study of vaccine response data One of many papers using ML to predict peptide or epitope interactions
4 Analytics	Woo et al, 2008 Dunn et al, 2017 Chandir et al, 2018	visual analytics GIS, visual analytics predictive analytics	Study the effects of sample stratification when datamining vaccine responses Determine impact of info exposure on HPV vaccine coverage Identify children at high risk of missing vaccinations

visualization or analytic platform. The table, and this study, classifies papers based on the technology highlighted by the author in terms of their perceived contributions.

The four layers of the Big Data stack are highlighted in Table 8 as a means of further illuminating the value of these tools in the development of vaccines. Those authors noted in the Data Layer have utilized some form of large dataset as a tool to leverage further research and spur communication and collaboration among vaccine researchers. Without these standardized datasets, the complexity of sharing data is greatly complicated because of differences in format, labeling, and simple access. This leads to the second layer, Integration and Ingestion. This layer could also include tools such as data warehousing and the like. While these tools are likely used by most of the researchers in the study, they were only noted by one who described the use of the SEND format when sharing data across institutions or projects. Using a standard data format removes many of the typical ETL tasks involved in making data from external sources useful to a project.

The use of various machine learning techniques is common among those researchers trying to predict certain biological aspects of vaccinology but a full comparison of the various

tools and their performance is beyond the scope of this paper. Even though some of the papers focused on comparing the performance of specific tools on a single task, the tasks varied across a wide spectrum of size and computations, making more global comparisons difficult. However, researchers such as Rosyda, Adji & Setiawan (2016) who use a neural network approach to predicting epitope activity on the P24 protein of the HIV virus provides an example of how these tools are employed. The amino acids within the epitope are first encoded and then used to create a training set for use in developing the neural network capabilities. These researchers found that multiple training sessions, known as cascade training, provided even better predictive performance and sensitivity.

The same can be said about the use of datamining tools in the performance of pharmacovigilance studies. These tools provide an important window into the performance of a vaccine after approval and implementation, but the specifics of each study differ widely. Another example of how datamining is put to use these tools to analyze data collected from throughout a human body to describe its response to a vaccination (Flanagan et al, 2013). These huge and complex datasets can be mined to allow researchers to understand on a much broader scale the human body's response to vaccinations.

The last category, Analytics, comprises all of the tools that researchers might use to analyze the results of their computation or try to draw information out of large databases, such as those described in the first layer of the stack. In most cases, the actual tool being used is not described fully, only the functionalities employed. In many cases, visual analytics are employed to understand patterns and relationships in the data or to enable interactions hidden from view to be highlighted in a useful manner. An illustration of this approach is the use of visual analytics and Geographical Information Systems (GIS) to determine the regional differences in vaccine acceptance based on social media data (Dunn et al, 2017). Using GIS tools, the data are not changed or transformed, but displayed in such a way that the underlying effects of location can be highlighted to enable easier understanding of the differences brought about by that single variable.

There is a large disparity between the number of journal articles and conference proceedings. This was surprising because many times, if not most, when information technologies are introduced into a new milieu, the initial forays into that area of inquiry happen at conferences. These conferences are where new ideas are tried out and then published in journals after they mature. In the present case, journal articles are much more numerous than conference papers. This might be an artifact of the manner in which the databases are searched or constructed, or possibly the nature of the industry in which these efforts take place. It would be interesting to see if other areas of literature in similar circumstances exhibit the same characteristics.

Another interesting characteristic of this body of literature is that, unlike other areas of technological inquiry, there are many more empirical papers than theoretical. Again, in other technological areas of inquiry, early literature is theoretical and/or prescriptive in nature, and then become more applied as the field matures (Kasten, 2020). In this case, there are many more empirical

uses of these technologies than prescriptive descriptions of systems yet to be built. Part of the change of balance has to do with the huge chunk of literature that deals specifically with analysis of adverse events surveillance. But, even when that literature is taken out, the large proportion of applied and empirical studies suggest that the researchers in this area have been better able to apply these technologies. Perhaps this represents a shortening of the learning curve as big data tools become more widely used and understood.

Very surprising is the lack of research surrounding the application of these technologies to the topics contained in the Miscellaneous theme such as supply chain management and manufacturing. It is hard to believe that the difficulties associated with transporting highly unstable materials such as vaccines, with their formidable temperature requirements, would not benefit from additional analytical tools and resulting information. The same might be said of the manufacturing function, which include very complicated processes and materials. The possibility exists that this research has taken place but is not narrowly identified as vaccine related literature, but rather more broadly in the realm of pharmaceuticals. Future research will explore this possibility.

Structured literature reviews have their inherent limitations. Choices of databases, search terms, or combinations thereof, can have significant effects on the results. This is coupled with the requirement that the researcher must exert judgement over which category or theme a particular document belongs to. Thus, differences in opinion over a particular document might result in slightly different results for a specific researcher. There is also the difficulty that classifying documents into closely related sub-themes can be very difficult because they deal with topics that are tightly interconnected with overlapping issues. For this study, each document was analyzed to determine what the most likely emphasis of the author is. There are many that could have found a home in more than one category and that is the very nature of interdisciplinary research. The difficulty in classification is exactly the reason for the value brought by this type of research.

CONCLUSION

The current study is intended to create an understanding in the reader of the uses to which Big Data tools such as analytics, datamining, machine learning and others within the vaccine development and distribution industry. The study also seeks to provide a framework for researchers to perform further research, especially in those areas which, as demonstrated by this study, have yet to be thoroughly investigated. Further research in this field is required to continue to illuminate the topic of Big Data technology in the vaccine industry specifically, and the pharmaceutical industry generally. While this would be helpful to researchers, it would also be helpful to practitioners and clinicians who are narrowly concerned with a specific aspect of the process and might not have a broader concept of the tools and technologies available to them.

REFERENCES

- Ackerman, M. E., Barouch, D. H., & Alter, G. (2017). Systems serology for evaluation of HIV vaccine trials. *Immunological Reviews*, 275(1), 262–270. doi:10.1111/immr.12503 PMID:28133810
- Agarwal, P., Tamer, M., Sahraei, H., & Budman, H. (2020). Deep learning for classification of profit-based operating regions in industrial processes. *Industrial & Engineering Chemistry Research*, 59(6), 2378–2395. doi:10.1021/acs.iecr.9b04737
- Altman, D. M. (2018). New tools for MHC research from machine learning and predictive algorithms to the tumour immunopeptidome. *Immunology*, 154(3), 329–330. doi:10.1111/imm.12956 PMID:29902342
- Banerjee, S., Basu, S., & Nasipuri, M. (2015). Big Data analytics and its prospects in computational proteomics. In *Information Systems Design and Intelligent Application* (pp. 591-598). Springer. doi:10.1007/978-81-322-2247-7_60
- Bell, A., Rich, A., Teng, M., Oreskovic, T., Bras, N. B., Mestrinho, L., Golubovic, S., Pristas, I., & Zejnilovic, L. (2019). Proactive advising: A machine learning driven approach to vaccine hesitancy. *IEEE International Conference on Healthcare Informatics*.
- Berendsen, M. L. T., Smits, J., Netea, M. G., & van der Ven, A. (2016). Non-specific effects of vaccines and stunting: Timing may be essential. *EBioMedicine*, 8, 341–348. doi:10.1016/j.ebiom.2016.05.010
- Bhattacharjee, P. S., Solanki, M., Bhattacharyya, R., Ehrenberg, I., & Sarma, S. (2015). VacSeen: A linked data-based information architecture to track vaccines using barcode scan authentication. SWAT4LS.
- Bhatti, U. A., Huang, M., Wang, H., Zhang, Y., Mehmood, A., & Di, W. (2018). Recommendation system for immunization coverage and monitoring. *Human Vaccines & Immunotherapeutics*, 14(1), 165–171. doi:10.1080/21645515.1379639
- Boehm, K. M., Bhinder, B., Raja, V. J., Dephoure, N., & Elemento, O. (2019). Predicting peptide presentation by major histocompatibility complex class I: An improved machine learning approach to the immunopeptidome. *BMC Bioinformatics*, 20(7), 1–11. doi:10.1186/s12859-018-2561-z PMID:30611210
- Bragazzi, N. L., Gianfredi, V., Villarini, M., Rosselli, R., Nasr, A., Hussein, A., Martini, M., & Behzadifar, M. (2018). Vaccines meet Big Data: State-of-the-art and future prospects. From the classical 3Is (“Isolate-Inactivate_Inject”) Vaccinology 1.0 to Vaccinology 3.0 Vaccinomics, and beyond: A historical overview. *Frontiers in Public Health*, 6(62), 1–9. doi:10.3389/fpubh.2018.00062 PMID:29556492
- Briggs, K. (2017). Making sense of SEND: The standard for exchange of nonclinical data. *Regulatory Toxicology and Pharmacology*, 91, 77–85. doi:10.1016/j.yrtph.2017.10.012 PMID:29066334
- Briner, R. B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. In D. M. Rousseau (Ed.), *The Oxford Handbook of Evidence-Based Management*. Oxford University Press.
- Cai, Y., Du, J., Huang, J., Ellenberg, S. S., Hennesy, S., Tao, C., & Chen, Y. (2017). A signal detection method for temporal variation of adverse effect with vaccine adverse event reporting system data. *BMC Medical Informatics and Decision Making*, 17(76), 93–100. doi:10.1186/s12911-017-0472-y PMID:28699543
- Chan, M.-S., Jamieson, K. H., & Albarracin, D. (2020). Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the United States. *Vaccine*, 38(40), 6236–6247. doi:10.1016/j.vaccine.2020.07.054 PMID:32792251
- Chandir, S., Siddiqi, D. A., Hussain, O. A., Naizi, T., Shah, M. T., Dharma, V. K., Habib, A., & Khan, A. J. (2018). Using predictive analytics to identify children at high risk of defaulting from a routine immunization program: Feasibility study. *JMIR Public Health and Surveillance*, 4(3), 1-12.
- Chaudhury, S., Duncan, E. H., Atre, T., Dutta, S., Spring, M. D., Leitner, W. W., & Bergmann-Leitner, E. S. (2020). Combining immunoprofiling with machine learning to assess the effects of adjuvant formulation on human vaccine-induced immunity. *Human Vaccines & Immunotherapeutics*, 16(2), 400–411. doi:10.1080/21645515.2019.1654807 PMID:31589550

- Chen, J., Chokshi, S., Hegde, R., Gonzalez, J., Iturrate, E., Aphinyanaphongs, Y., & Mann, D. (2020). Development, implementation, and evaluation of a personalized machine learning algorithm for clinical decision support: Case study with shingles vaccination. *Journal of Medical Internet Research*, 22(4), 1–12. doi:10.2196/16848 PMID:32347813
- Chuan, C.-H. (2018). Classifying eligibility criteria in clinical trials using active deep learning. *17th IEEE International Conference on Machine Learning and Applications*.
- Clark, K., Cavicchi, J., Jensen, K., Fitzgerald, R., Bennett, A., & Kowalski, S. P. (2011). Patent data mining: A tool for accelerating HIV vaccine innovation. *Vaccine*, 29(24), 4086–4093. doi:10.1016/j.vaccine.2011.03.052
- Conti, S., & Karplus, M. (2019). Estimation of the breadth of CD4bs targeting HIV antibodies by molecular modeling and machine learning. *PLoS Computational Biology*, 15(4), 1–22. 10.1371/journal.pcbi.1006954
- Cotugno, N., De Armas, L., Pallikkuth, S., Rossi, P., Palma, P., & Pahwa, S. (2015). Paediatric HIV infection in the ‘omics era: Defining transcriptional signatures of viral control and vaccine responses. *Journal of Virus Eradication*, 1(3), 153–158. doi:10.1016/S2055-6640(20)30507-0 PMID:26807446
- Cotugno, N., Santilli, V., Pascucci, G. R., Manno, E. C., De Armas, L., Pallikkuth, S., Deodati, A., Amodio, D., Zangari, P., Zicari, S., Ruggiero, A., Fortin, M., Bromley, C., Pahwa, R., Rossi, P., Pahwa, S., & Palma, P. (2020). Artificial intelligence applied to *in vitro* gene expression testing (IVIGET) to predict trivalent inactivated influenza vaccine immunogenicity in HIV infected children. *Frontiers in Immunology*, 11, 1–13. doi:10.3389/fimmu.2020.559590 PMID:33123133
- Dalsass, M., Brozzi, A., Medini, D., & Rappuoli, R. (2019). Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery. *Frontiers in Immunology*, 10, 1–12. Advance online publication. doi:10.3389/fimmu.2019.00113 PMID:30837982
- de Bie, S., Verhamme, K. M. C., Straus, S. M. J. M., Stricker, B. H. C., & Sturkenboom, M. C. J. M. (2012). Vaccine-based subgroup analysis in VigiBase. *Drug Safety*, 35(4), 335–346. doi:10.2165/11598120-000000000-00000 PMID:22435344
- de Sousa, T. N., & de Menezes Neto, A. (2013). “Omics” in the study of the major parasitic diseases malaria and schistosomiasis. *Infection, Genetics and Evolution*, 19, 258–273. doi:10.1016/j.meegid.2013.07.008 PMID:23871773
- Dehoot, A. M., Chirove, F., & Ndifon, W. (2018). Trans-allelic model for prediction of peptide: MHC-11 interactions. *Frontiers in Immunology*, 9, 1–11. Article, 1410. Advance online publication. doi:10.3389/fimmu.2018.01410 PMID:29988560
- Dharayani, R., Wibowo, W. C., Ruldeviyani, Y., & Gandhi, A. (2019). *Genomic anomaly searching with BLAST algorithm using MapReduce framework in big data platform*. IWBIS.
- Du, J., Xu, J., Song, H., & Tao, C. (2017). Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Medical Informatics and Decision Making*, 17(69), 63–70. doi:10.1186/s12911-017-0469-6
- Dunn, A. G., Surian, D., Leask, J., Dey, A., Mandl, K. D., & Coiera, E. (2017). Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine*, 35(23), 3033–3040. doi:10.1016/j.vaccine.2017.04.060 PMID:28461067
- El-Manzalaway, Y., Munoz, E. E., Lidner, S. E., & Honavar, V. (2016). PlasmOSEP: Predicting surface-exposed proteins on the malaria parasite using semisupervised self-training and expert-annotated data. *Proteomics*, 16(23), 2967–2976. doi:10.1002/pmic.201600249 PMID:27714937
- Flanagan, K. L., Noho-Konteh, F., Ghazal, P., & Dickinson, P. (2013). Transcriptional profiling technology for studying vaccine responses: An untapped goldmine. *Methods (San Diego, Calif.)*, 60(3), 269–274. doi:10.1016/j.ymeth.2013.03.032 PMID:23578546
- Frampton, M. (2018). *Complete Guide to Open Source Big Data Stack*. Springer. doi:10.1007/978-1-4842-2149-5
- Getman, R., Helmi, M., Roberts, H., Yansane, A., Cutler, D., & Seymour, B. (2018). Vaccine hesitancy and online information: The influence of digital networks. *Health Education & Behavior*, 45(4), 599–606. doi:10.1177/1090198117739673 PMID:29267129

- Gilbert, P. B., & Luedtke, A. R. (2018). Statistical learning methods to determine immune correlates of herpes zoster in vaccine efficacy trials. *The Journal of Infectious Diseases*, 218(suppl_2), s99–s101. doi:10.1093/infdis/jiy421 PMID:30247601
- Haber, P., Moro, P. L., Ng, C., Dores, G. M., Perez-Vilar, S., Marquez, P. L., & Cano, M. (2020). Safety review of tetanus toxoid, reduced diphtheria toxoid, acellular pertussis vaccines (Tdao) in adults aged ≥ 65 years, Vaccine Adverse Event Reporting System (VAERS), United States, September 2010–December 2018. *Vaccine*, 38(6), 1476–1480. doi:10.1016/j.vaccine.2019.11.074 PMID:31883809
- Haber, P., Moro, P. L., Ng, C., Doros, G. M., Lewis, P., & Cano, M. (2019). Post licensure surveillance of trivalent adjuvanted influenza vaccine (allV3;Fluad), vaccine Adverse Event Reporting System (VAERS), United States, July 2016–June 2018. *Vaccine*, 37(11), 1516–1520. doi:10.1016/j.vaccine.2019.01.052 PMID:30739795
- He, Y. (2014). Ontology-supported research on vaccine efficacy, safety and integrative biological networks. *Expert Review of Vaccines*, 13(7), 825–841. doi:10.1586/14760584.2014.923762 PMID:24909153
- Heinson, A. I., Ewing, R. M., Holloway, J. W., Woelk, C. H., & Niranjani, M. (2019). An evaluation of different classification algorithms for protein sequence-based reverse vaccinology prediction. *PLoS One*, 14(12), 1–13. doi:10.1371/journal.pone.0226256 PMID:31834914
- Hemedan, A., Elaziz, M. A., Jiao, P., Alavi, A. H., Bahgat, M., Ostaszewski, M., Schneider, R., Ghazy, H. A., Ewees, A. A., & Lu, S. (2020). Prediction of the vaccine-derived Poliovirus outbreak incidence: A hybrid machine learning approach. *Scientific Reports*, 10(5058), 1–12. doi:10.1038/s41598-020-61853-y PMID:32193487
- Henderson, L. (2016). Innovations in patient matching. *Applied Clinical Trials*, 25(8/9), 1.
- Henneges, C., Huster, S., & Zell, A. (2009). An artificial T cell immune system for predicting MHC-II binding peptides. *IEEE Symposium on Artificial Life*.
- Hovav, S., Tell, H., Levner, E., Ptuskin, A., & Herbon, A. (2017). Health care analytics and Big Data management in influenza vaccination programs: Use of information-entropy approach. In *The Analytics Process* (pp. 211–237). doi:10.1109/ALIFE.2009.4937708
- Iqbal, S., Shi, J., Seib, K., Lewis, P., Moro, P. L., Woo, E. J., Shimabukuro, T., & Orenstein, W. A. (2015). Preparation for global introduction of inactivated poliovirus vaccine: Safety evidence from the US Vaccine Adverse Event Reporting System, 2000–12. *The Lancet. Infectious Diseases*, 15(10), 1175–1182. doi:10.1016/S1473-3099(15)00059-6 PMID:26289956
- Izak, D., Klim, J., & Kaczanowski, S. (2018). Host-parasite interactions and ecology of the malaria parasite—a bioinformatics approach. *Briefings in Functional Genomics*, 17(6), 451–457. doi:10.1093/bfpg/ely013 PMID:29697785
- Jordan, K., Dossou, P.-E., & Chang, J. Jr. (2019). Using lean manufacturing and machine learning for improving medicines procurement and dispatching in a hospital. *Procedia Manufacturing*, 38, 1034–1041. doi:10.1016/j.promfg.2020.01.189
- Kagashe, I., Yan, Z., & Suheryani, I. (2017). Enhancing seasonal influenza surveillance: Topic analysis of widely used medicinal drugs using Twitter data. *Journal of Medical Internet Research*, 19(9), 1–14. doi:10.2196/jmir.7393 PMID:28899847
- Kar, P., Ruiz-Perez, L., Arooj, M., & Mancera, R. L. (2018). Current methods for the prediction of T-cell epitopes. *Peptide Science (Hoboken, N.J.)*, 110(2), 1–17. doi:10.1002/pep2.24046
- Kasten, J. (2020). Big data applications in healthcare administration. *International Journal of Big Data Applications in Healthcare*, 5(2), 12–37.
- Khanna, D., & Rana, P. S. (2017). Multilevel ensemble model for prediction of IgA and IgG antibodies. *Immunology Letters*, 184, 51–60. doi:10.1016/j.imlet.2017.01.017 PMID:28214535
- Khanna, D., & Rana, P. S. (2020). Improvement in prediction of antigenic epitopes using stacked generalisation: An ensemble approach. *IEEE Systems Biology*, 14(1), 1–7. doi:10.1049/iet-syb.2018.5083 PMID:31931475

- Kim, S., Kim, H. S., Kim, E., Lee, M. G., Shin, E.-C., Paik, S., & Kim, S. (2018). Neopepsee: Accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 29(4), 1030–1036. doi:10.1093/annonc/mdy022 PMID:29360924
- Kim, Y., Lee, J., Ha, K., Lee, W.-K., Heo, D. R., Woo, J. R., & Yu, S. J. (2019). A computational framework for deep learning-based epitope prediction by using structure and sequence information. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.
- Krittana Wong, C., Tunhasiriwet, A., Chirapongsathorn, S., & Kitai, T. (2017). Tweeting influenza vaccine to cardiovascular health community. *European Journal of Cardiovascular Nursing*, 16(8), 704–706. doi:10.1177/1474515117707867
- Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. *Information Sciences*, 467, 312–322. doi:10.1016/j.ins.2018.08.019
- Kwofie, S. K., Broni, E., Teye, J., Quansah, E., Issah, I., Wilson, M. D., Miller, W. A. III, Tiburu, E. K., & Bonney, J. H. K. (2019). Pharmacoinformatics-based identification of potential bioactive compounds against Ebola virus protein VP24. *Computers in Biology and Medicine*, 113, 1–18. doi:10.1016/j.combiomed.2019.103414 PMID:31536833
- Lalmuanawma, S., Hussain, J., & Chhakhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons, and Fractals*, 139, 1–6. doi:10.1016/j.chaos.2020.110059 PMID:32834612
- Lee, A. J., Das, S. R., Wang, W., Fitzgerald, T., Pickett, B. E., Aevermann, B. D., Topham, D. J., Falsey, A. R., & Scheuermann, R. H. (2015). Diversifying selection analysis predicts antigenic evolution of 2009 pandemic H1N1 influenza A virus in humans. *Journal of Virology*, 89(10), 5427–5440. doi:10.1128/JVI.03636-14 PMID:25741011
- Liu, G., Carter, B., Bricken, T., Jain, S., Viard, M., Carrington, M., & Gifford, D. K. (2020). Computationally optimized SARS-CoV-2 MHC class I and II vaccine formulations predicted to target human haplotype distributions. *Cell Systems*, 11(2), 131–144. doi:10.1016/j.cels.2020.06.009 PMID:32721383
- Liu, K., Li, S., Qian, Z. M., Dharmage, S. C., Bloom, M. S., Heinrich, J., Jalaludin, B., Markevych, I., Morawska, L., Knibbs, L. D., Hinyard, L., Xian, H., Liu, S., Line, S., Leskinen, A., Komppula, M., Jalava, P., Roponen, M., Hu, L.-W., Zeng, X.-W., Hu, W., Chen, G., Yang, B.-Y., Guo, Y., & Dong, G.-H. (n.d.). Benefits of influenza vaccination on the associations between ambient air pollution and allergic respiratory diseases in children and adolescents: New insights from the Seven Northeastern Cities study in China. *Environmental Pollution*, 256, 1–10. 10.1016/j.envpol.2019.113434
- Lopalco, P. L., & DeStefano, F. (2015). The complementary roles of phase 3 trials and post-licensure surveillance in the evaluation of new vaccines. *Vaccine*, 33(13), 1541–1548. doi:10.1016/j.vaccine.2014.10.047 PMID:25444788
- Lu, X., Shi, H., Biswas, R., Javed, M. H., & Panda, D. K. (2018). DLoBD: A comprehensive study of deep learning of big data stacks on HPC clusters. *IEEE Transactions on Multi-Scale Computing Systems*, 4(4), 635–648. doi:10.1109/TMSCS.2018.2845886
- Martin, S., Kilich, E., Dada, S., Kummervold, P. E., Denny, C., Paterson, P., & Larsen, H. J. (2020). Vaccines for pregnant women...?!-Mapping maternal vaccination discourse and stance on social media over six months. *Vaccine*, 38(42), 6627–6637. doi:10.1016/j.vaccine.2020.07.072 PMID:32788136
- Massey, P. M., Leader, A., Yom-Tov, E., Budenz, A., Fisher, K., & Klassen, A. C. (2016). Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. *Journal of Medical Internet Research*, 18(12), 1–11. doi:10.2196/jmir.6670 PMID:27919863
- Meyer, S. B., Violette, R., Aggarwal, R., Simeoni, M., MacDougall, H., & Waite, N. (2019). Vaccine hesitancy and Web 2.0: Exploring how attitudes and beliefs about influenza vaccination are exchanged in online threaded user comments. *Vaccine*, 37(13), 1769–1774. doi:10.1016/j.vaccine.2019.02.028 PMID:30826142
- Miranda, P. O., Cubitt, B., Jacob, N. T., Janda, K. D., & de la Torre, J. C. (2018). Mining a Krohnke Pyridine Library for anti-arenavirus activity. *ACS Infectious Diseases*, 4(5), 815–824. doi:10.1021/acsinfecdis.7b00236 PMID:29405696

Montagna, M. T., De Giglio, O., Napoli, C., Fasano, F., Diella, G., Donnoli, R., Caggiano, G., Tafuri, S., Lopalco, P. L., & Agodi, A. (2020). Adherence to vaccination policy among public health professionals: Results of national survey in Italy. *Vaccines*, *8*(379), 1–16. doi:10.3390/vaccines8030379 PMID:32664507

Morgan, C. (2019). Analytics and metrics help pinpoint costs of study start-up. *Applied Clinical Trials*, *28*(1/2), 10–15.

Munteanu, C. R., Pedreira, N., Dorado, J., Pazos, A., Perez-Montoto, L. G., Ubeira, F. M., & Gonzalez-Diaz, H. (2014). LECTINPred: Web server that uses complex networks of protein structure for prediction of lectins with potential use as cancer biomarkers or in parasite vaccine design. *Molecular Informatics*, *33*(4), 276–285. doi:10.1002/minf.201300027 PMID:27485774

Myers, T. R., McNeil, M. M., Ng, C. S., Li, R., Lewis, P. W., & Cano, M. (2017). Adverse events following quadrivalent meningococcal CRM-conjugate vaccine (Menveo) reported to the Vaccine Adverse Event Reporting System (VAERS), 2010–2015. *Vaccine*, *35*(14), 1758–1763. doi:10.1016/j.vaccine.2017.02.030 PMID:28262331

Nagpal, G., Chaudhary, K., Agrawal, P., & Raghava, P. S. (2018). Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. *Journal of Translational Medicine*, *16*(181), 1–15. doi:10.1186/s12967-018-1560-1 PMID:29970096

Nanni, L., & Lumini, A. (2009). An ensemble of support vector machines for predicting virulent proteins. *Expert Systems with Applications*, *36*(4), 7458–7462. doi:10.1016/j.eswa.2008.09.036

Nawa, N., Kogaki, S., Takahashi, K., Ishida, H., Baden, H., Katsuragi, S., Narita, J., Tanaka-Taya, K., & Ozono, K. (2016). Analysis of public concerns about influenza vaccinations by mining a massive online question dataset in Japan. *Vaccine*, *34*(27), 3207–3213. doi:10.1016/j.vaccine.2016.01.008 PMID:26776467

Oberg, A. L., McKinney, B. A., Schaid, D. J., Pankratz, V. S., Kennedy, R. B., & Poland, G. A. (2015). Lessons learned in the analysis of high-dimensional data in vaccinomics. *Vaccine*, *33*(40), 5262–5270. doi:10.1016/j.vaccine.2015.04.088 PMID:25957070

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future – big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, *375*(13), 1216–1219. doi:10.1056/NEJMp1606181 PMID:27682033

Okuhara, T., Ishikawa, H., Okada, M., Kato, M., & Kiuchi, T. (2018). Contents of Japanese pro- and anti-HPV vaccination websites: A text mining analysis. *Patient Education and Counseling*, *101*(3), 406–413. doi:10.1016/j.pec.2017.09.014 PMID:29031425

Okuhara, T., Ishikawa, H., Okada, M., Kato, M., & Kiuchi, T. (2019a). Japanese anti-versus pro-influenza vaccination websites: A text-mining analysis. *Health Promotion International*, *34*(3), 552–566. doi:10.1093/heapro/day015 PMID:29584863

Okuhara, T., Ishikawa, H., Okada, M., Kato, M., & Kiuchi, T. (2019b). Newspaper coverage before and after the HPV vaccination crisis began in Japan: A text mining analysis. *BMC Public Health*, *19*(770), 1–15. doi:10.1186/s12889-019-7097-2 PMID:31208394

Olafsdottir, T., Lindqvist, M., & Harandi, A. M. (2015). Molecular signatures of vaccine adjuvants. *Vaccine*, *33*(40), 5302–5307. doi:10.1016/j.vaccine.2015.04.099 PMID:25989447

Ong, E., Wong, M. U., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Frontiers in Immunology*, *11*, 1–13. Advance online publication. doi:10.3389/fimmu.2020.01581 PMID:32719684

Panosos, A. D., Bury, T. M., Wang, C., Schonfeld, J., Mohanty, S. P., Nyhan, B., Salathe, M., & Bauch, C. T. (2017). Critical dynamics in population vaccinating behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(52), 13762–13767. Advance online publication. doi:10.1073/pnas.1704093114 PMID:29229821

Parvande, S., Poland, G. A., Kennedy, R. B., & McKinney, B. A. (2019). Multi-level model to predict antibody response to influenza vaccine using gene expression interaction network feature selection. *Microorganisms*, *7*(79), 1–17. doi:10.3390/microorganisms7030079 PMID:30875727

- Pennisi, M., Russo, G., & Pappalardo, F. (2018). Combining parallel genetic algorithms and machine learning to improve the research of optimal vaccination protocols. *26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*.
- Pittala, S., Morrison, K. S., & Akerman, M. E. (2019). Systems serology for decoding infection and vaccine-induced antibody responses to HIV-1. *Current Opinions HIV AIDS, 14*, 253-264. doi:10.1109/PDP2018.2018.00070
- Rodrigo, C., & Luciani, F. (2019). Dynamic interactions between RNA viruses and human hosts unravelled by a decade of next generation sequencing. *BBA-General Subjects, 1863*(2), 511–519. doi:10.1016/j.bbagen.2018.12.003 PMID:30528489
- Rosyda, M., Adji, T. B., & Setiawan, N. A. (2016). Cascade neural network for predicting epitope on P24 HIV-1. *Proceedings of the 1st International Conference on Science and Technology 2015 (ICST-2015)*. doi:10.1063/1.4958500
- Sakr, S. (2017). Big data processing stacks. *IT Professional, 19*(1), 34–41. doi:10.1109/MITP.2017.6
- Sher, G., Zhi, D., & Zhang, S. (2017). DRREP: Deep ridge regressed epitope predictor. *BMC Genomics, 18*(S6), 55–65. doi:10.1186/s12864-017-4024-8 PMID:28984193
- Singh, D., Singh, P., & Sisodia, D. S. (2019). Compositional model based on factorial evolution for realizing multi-task learning in bacterial virulent protein prediction. *Artificial Intelligence in Medicine, 101*, 1–6. doi:10.1016/j.artmed.2019.101757 PMID:31813491
- Smith, C. C., Chai, S., Washington, A. R., Lee, S. J., Landoni, E., Field, K., Garness, J., Bixby, L. M., Selitsky, S. R., Parker, J. S., Savoldo, B., Serody, J. S., & Vincent, B. G. (2019). Machine-learning prediction of tumor antigen immunogenicity in the selection of therapeutic epitopes. *Cancer Immunology Research, 7*(10), 1591–1604. doi:10.1158/2326-6066.CIR-19-0155 PMID:31515258
- Solihah, B., Azhari, A., & Musdholifah, A. (2020). Enhancement of conformational B-cell epitope prediction using CluSMOTE. *Peer J. Computer Science, 6*, 1-17. 10.7717/peerj-cs.275
- Sujaree, K. (2019). Designing a vaccine cold chain network in northern Thailand using the max-min ant system. *Journal of Science and Technology, 26*(3), 257–265.
- Svanström, H., Callreus, T., & Hviid, A. (2010). Temporal data mining for adverse events following immunization in nationwide Danish healthcare databases. *Drug Safety, 33*(11), 1015–1025. doi:10.2165/11537630-000000000-00000 PMID:20925439
- Tavoschi, L., Quattrone, F., D'Andrea, E., Ducange, P., Vabanesi, M., Marcelloni, F., & Lopalco, P. L. (2020). Twitter as a sentinel tool to monitor public opinion on vaccination: An opinion mining analysis from September 2016 to August 2017 in Italy. *Human Vaccines & Immunotherapeutics, 16*(5), 1062–1069. doi:10.1080/21645515.2020.1714311 PMID:32118519
- Tomeny, T. S., Vargo, C. J., & El-Toukhy, S. (2017). Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. *Social Science & Medicine, 191*, 168–175. doi:10.1016/j.socscimed.2017.08.041 PMID:28926775
- Tomic, A., Tomic, I., Dekker, C. L., Maecker, H. T., & Davis, M. M. (2019). The FluPRINT dataset, a multidimensional analysis of the influenza vaccine imprint on the immune system. *Scientific Data, 6*(1), 1–11. doi:10.1038/s41597-019-0213-4 PMID:31636302
- Tomic, A., Tomic, I., Rosenberg-Hasson, Y., Dekker, C. L., Maecker, H. T., & Davis, M. M. (2019). SIMON, an automated machine learning system, reveals immune signatures of influenza vaccine responses. *Journal of Immunology (Baltimore, Md.: 1950), 203*(3), 749–759. doi:10.4049/jimmunol.1900033 PMID:31201239
- Trifiro, G., Sultana, J., & Bate, A. (2018). From Big Data to smart data for pharmacovigilance: The role of healthcare databases and other emerging sources. *Drug Safety, 41*(2), 143–149. doi:10.1007/s40264-017-0592-4 PMID:28840504
- Tumiotto, C., Riviere, L., Bellacave, P., Recordon-Pinson, P., Vilain-Parce, A., Guidicelli, G.-L., & Fleury, H. (2017). Sanger and next-generation sequencing data for characterization of CTL epitopes in archived HIV-1 proviral DNA. *PLoS One, 12*(9), 1–10. doi:10.1371/journal.pone.0185211 PMID:28934310

- Vaishya, R., Javaid, M., Khan, I. H., & Maleem, A. (2020). Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome, 14*(4), 337–339. doi:10.1016/j.dsx.2020.04.012 PMID:32305024
- Vang, Y. S., & Xie, X. (2017). HLA class I binding prediction via convolutional neural networks. *Bioinformatics (Oxford, England), 33*(17), 2658–2665. doi:10.1093/bioinformatics/btx264 PMID:28444127
- Wang, S. V., Maro, J. C., Baro, E., Izem, R., Dashevsky, I., Rogers, J. R., Nguyen, M., Gagne, J. J., Patorno, E., Huybrechts, K. F., Major, J. M., Zhou, E., Reidy, M., Cosgrove, A., Schneeweiss, S., & Kulldorff, M. (2018, November). Data mining for adverse drug events with a propensity score-matched tree-based scan statistic. *Epidemiology (Cambridge, Mass.), 29*(6), 895–903. doi:10.1097/EDE.0000000000000907 PMID:30074538
- Woo, E. J., Ball, R., Burwen, D. R., & Braun, M. M. (2008). Effects of stratification on data mining in the US Vaccine Adverse Effect Reporting System (VAERS). *Drug Safety, 31*(8), 667–674. doi:10.2165/00002018-200831080-00003 PMID:18636785
- Wu, H., Mu, J., Da, T., Xu, M., Taylor, R. H., Iordachita, I., & Chirikjian, G. S. (2019). Multi-Mosquito object detection and 2D pose estimation for automation of PfSPZ malaria vaccine production. IEEE 15th International Conference on Automation Science and Engineering (CASE).
- Yildirim, M., Keskinocak, P., Pelton, S., Pickering, L., & Yildirim, I. (2020). Who is at risk of 13-valent conjugated pneumococcal vaccine failure? *Vaccine, 38*, 1671–1677. doi:10.1016/j.vaccine.2019.10.060
- Yong, B., Shen, J., Liu, X., Li, F., Chen, H., & Zhou, Q. (2020). An intelligent blockchain-based system for safe vaccine supply and supervision. *International Journal of Information Management, 52*, 1–12. doi:10.1016/j.ijinfomgt.2019.10.009
- Young, J. A., Fivelman, Q. L., Blair, P. L., de la Vega, P., Le Roch, K. G., Zhou, Y., Carucci, D. J., Baker, D. A., & Winzeler, E. A. (2005). The *Plasmodium falciparum* sexual development transcriptome: A microarray analysis using ontology-based pattern identification. *Molecular and Biochemical Parasitology, 143*(1), 67–79. doi:10.1016/j.molbiopara.2005.05.007 PMID:16005087
- Yu, H., Wang, G., Zhao, Z., Wang, H., & Wang, Z. (2019). Chicken embryo fertility detection based on PPG and convolutional neural network. *Infrared Physics & Technology, 103*, 1–6. doi:10.1016/j.infrared.2019.103075
- Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Social Media + Society, 5*(3), 1–12. doi:10.1177/2056305119865465
- Zame, W. R., Bica, I., Shen, C., Curth, A., Lee, H.-S., Bailey, S., Weatherall, J., Wright, D., Bretz, F., & van der Schaar, M. (2020). Machine learning for clinical trials in the era of COVID-19. *Statistics in Biopharmaceutical Research, 12*(4), 506–517. doi:10.1080/19466315.2020.1797867
- Zhang, G., Lin, H. H., Keskin, D. B., Reinherz, E. L., & Brusica, V. (2011). Dana-Farber repository for machine learning in immunology. *Journal of Immunological Methods, 374*(1-2), 18–25. doi:10.1016/j.jim.2011.07.007 PMID:21782820
- Zhang, W., & Niu, Y. (2010). Predicting flexible length linear B-cell epitopes using pairwise sequence similarity. *3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010)*.
- Zhang, Y., Aevermann, B., Anderson, T. K., Burke, D. F., Dauphin, G., Gu, Z., He, S., Kumar, S., Larsen, C. N., Lee, A. J., Li, X., Macken, C., Mahaffey, C., Pickett, B. E., Reardon, B., Smith, T., Stewart, L., Suloway, C., Sun, G., & Scheuermann, R. H. et al. (2017). Influenza research database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Research, 45*(D1), D466–D474. doi:10.1093/nar/gkw857 PMID:27679478
- Zung, J. (2019). The growing importance of real-time access to clinical study performance. *Applied Clinical Trials, 28*(11), 7.