

## The Application of Data Science Techniques and Algorithms in Women's Health Studies

Ikpe Justice Akpan, BSc, MSc, Ph.D.<sup>\*1,3</sup> and Obianuju Genevieve Aguolu, MBBS, MPH, Ph.D.<sup>2</sup>

1. Department of Management & Information Systems, Kent State University, 330 University Drive NE, New Philadelphia, OH 44663, USA. [iakpan@kent.edu](mailto:iakpan@kent.edu) (*\*Corresponding Author*)
2. Infectious Disease Internal Medicine Department, Yale School of Medicine, Yale University, New Haven, Connecticut, USA.
3. Ibom International Center for Research and Scholarship, 107-9618 Tecumseh Road East, Windsor ON, N8B 1A2, Canada.

### Abstract

**Objective:** Evaluate and map data science methods employed to solve health conditions of women, examine the problems tackled and the effectiveness.

**Research Method:** Text analytics, science mapping, and descriptive evaluation of data science methods utilized in women-related health research.

**Findings:** (i). The trends in scholarships using data science methods indicate gaps between women and men relating to health burden and access to health. (ii). The coronavirus (SARS-CoV-2) outbreak and the ongoing COVID-19 pandemic tend to widen the identified health gaps, increasing the disease burden for women, while reducing access to health. There are noticeable additional health burdens on pregnant women and those with several health conditions (breast cancer, gynecologic oncology, cardiovascular disease, and more). (iii). Over 95% of studies using data science methods (artificial intelligence, machine learning, novel algorithms, predictive, big data, visual analytics, clinical decision support systems, or a combination of the methods) indicate significant effectiveness. (iv). Mapping of the scientific literature to authors, sources, and countries show an upward trend; 997 (16%), 113 (1.33%), and 57 (2.63%) per article, respectively. About 95% of research utilizing data science methods in women's health studies occurred within the last four (4) years.

**Conclusions:** The application of data science methods in tackling different health problems of women is effective and growing, and capable of easing the burden of health in women. The ongoing COVID-19 pandemic tends to compound the health burden for women more than men. Policymakers must do more to improve access to health for women.

**Keywords:** data science methods, machine learning, algorithms, big data, visual analytics, women's health

## 1. INTRODUCTION

The use of information systems, computer science methods, and operations research techniques to improve operations performance in the healthcare sector has a long history.<sup>1-5</sup> Data science (DS) techniques and analytical methods are also employed to mitigate health disparities across social, cultural, and sex classification more recently.<sup>6,7</sup> The goal is to eliminate the known gaps in healthcare access, reduce disease burden, and improve clinical outcomes and healthcare for all.<sup>8,9</sup>

The health gaps between men and women in terms of disease burden and access to health are well established in the social science and healthcare literature, including the ongoing coronavirus disease (COVID-19).<sup>122</sup> Gaps are also reported during the ongoing coronavirus disease (COVID-19) pandemic regarding disease burden and access to health. While the pandemic places an additional health burden on women than men, such as pregnant women and those with pre-existing breast cancer, gynecologic oncology, and other health conditions, fewer women (on average) have access to health compared to men. For example, in studies reported in this paper, pregnant women diagnosed with COVID-19 can have severe morbidities and have greater mortality risk. Simultaneously, child-care responsibility and economic burdens rests on working mothers during lockdowns and working from home rests more.<sup>45,74,100,103,122</sup>

The recent advances in computer processing power and the availability of massive and complex data comes the use of DS tools and methods and big data analytics techniques to conduct healthcare and biomedical sciences research and practice.<sup>10,11</sup> The ongoing advances in DS and analytics methods include machine learning, deep learning, artificial intelligence, which in addition to visual analytics and information visualization, enhances knowledge discovery, sensemaking, understanding, and offering solutions to complex medical and health problems.<sup>11-15,108</sup> Recent studies identify healthcare research

and practice as one of the active fields utilizing data science technologies and methods to enhance clinical decision support systems and improving health outcomes.<sup>5,10,11,60</sup> Therefore, this study undertakes an evaluation of the data science techniques and novel algorithms employed in women's health studies. We also map specific approaches to the type of problems solved, identify what works and what does not, and identify areas that require more work.

## **2. THEORETICAL BACKGROUND**

This section examines the peculiarity of women's health and the disparity in terms of the burden and access to health.

### **2.1 Women's Health**

Certain diseases and conditions primarily affect women due to differences in their body anatomy and physiology. Some peculiar health conditions that impact women include breast health, obstetric and gynecological issues. Others are and biocompatibility issues related to implants used in women.<sup>17</sup> Furthermore, several factors can influence the way women respond to common diseases and medical products. These factors include intrinsic (age, genetics, hormones, body size, sex-specific physiology, pregnancy), extrinsic factors (diet, sociocultural issues, environment).

Sex-specific biomarkers vary in females by age because of variations in hormone level, co-morbidities, and taking multiple medications. The toxicity of some medications can increase due to changes in a woman's physiology during pregnancy. Also, certain medications can be contraindicated in women during lactation due to potential toxicity or teratogenicity to the growing fetus, limiting their therapy choices. Some common diseases that may affect women differently include osteoporosis,

autoimmune diseases, certain cancers such as breast, uterine, cervical cancers, cardiovascular diseases, lung diseases, neurological diseases, and certain psychiatric conditions.

Osteoporosis is one of the common health problems experienced by older women due to their physiology. Women experience fluctuations in estrogen throughout life. Depletion of estrogen level causes osteoporosis, a risk factor for fractures, increased morbidity, and mortality. Moreover, this condition negatively impacts postmenopausal women's physical, emotional, and mental well-being.<sup>18</sup> Thus, it is essential to consider women's risk factors when providing care or managing cases.

The disparity in Women's health is also prevalent in autoimmune diseases. Those conditions include autoimmune thyroid, multiple sclerosis, and rheumatological systemic autoimmune. Women also carry more disease burden on systemic lupus erythematosus, Sjögren's syndrome effect, and rheumatoid arthritis, and represent most of these cases.<sup>8,19,45</sup> The reasons for this imbalance are unclear but attributed to genetic (X-linked) and hormonal factors. Furthermore, the women's health imbalance is present in the course and prognosis of the diseases. More studies are needed to investigate if this imbalance is due to differences in the disease's biology, how men and women respond to therapy or the treatment.<sup>8,9,19,45</sup>

Estrogen has anti-inflammatory effects. The fall in estrogen levels as women age predisposes them to arthritis. Older women are more likely to get arthritis than men of the same age. They are also more likely to experience worse pain from it more than men. They are also more vulnerable to rheumatoid arthritis, a highly debilitating form of arthritis. Furthermore, arthritis tends to affect women's limb joints more than men due to pregnancy changes and female anatomy.

In the United States, cardiovascular diseases are the leading cause of death among women. Most of these deaths are caused by coronary heart diseases and are usually sudden. Coronary heart diseases are responsible for 20% of female deaths. About 1 in 16 women aged 20 and older (6.2%) have

coronary heart disease, the most common heart disease.<sup>20</sup> Framingham's study, pioneering research in the sex differences, demonstrated that women have greater primary risk factors for heart diseases than men.<sup>21</sup> Furthermore, studies found that women are more likely to have worse outcomes from heart diseases. Possible factors related to worse effects in women include delays in seeking medical care or voluntary discharge. The reasons for these delays are unknown. More studies should be directed at understanding these risk factors and should be made available to the public, particularly women.

There are sex-specific disparities in the public health burden of cancers due to variations at the genetic/molecular level and sex hormones such as estrogen.<sup>22</sup> The genetic and hormonal differences influence gene expressions of certain cancers, drug metabolism, and therefore the effect of chemotherapy. These differences should be considered in cancer research to avoid disparity of chemotherapy's efficacy and toxicity between sexes. Women are more likely to experience psychiatric disorders, alcohol abuse, substance use disorders, conduct disorder, syndromal adulthood antisocial behavior without conduct disorder,<sup>23</sup> tranquilizer abuse with social and specific phobias,<sup>24</sup> and major depressive disorder and anxiety disorders compared to men.<sup>26</sup> Females are more likely to experience the double stigma of being both alcoholic and morally degenerate. They are more likely to face societal stigmas at work, home, school, community. Furthermore, women with mental health disorders and drug addictions experience more stigma than men.<sup>25</sup> The biological difference is attributed to the sex chromosomes. Awareness and understanding of the sexual dimorphism in neuropsychiatric disorders, increased risk of neurodevelopmental and psychiatric disorders in sex-linked genetic disorders can guide the management of these disorders.<sup>26</sup>

## **2.2 Disparities in Women's Access to Healthcare**

Previous studies have identified sex disparities in access to care.<sup>7,8,27</sup> While health needs are significantly greater among older women than men of the same age, women have fewer economic resources.<sup>27</sup> Even

after controlling for health needs, there are still sex differences in preventive care and increased differences in hospital services. Women are less likely to have hospital stays or outpatient services<sup>28</sup> than men with similar demographic and health profiles. Women are more likely than men to report insufficient money for medications, delays in care,<sup>29</sup> insufficient money for health care, and insufficient money for mental health care. Studies show that while women often develop drug and alcohol addictions faster than men, they are likely to face multiple barriers to substance use treatment and are less likely to see treatment.<sup>25</sup> They are more likely to have poorer treatment outcomes because of more inadequate resources. Some factors that may contribute to a delay in women seeking psychiatric care include hesitancy to discuss reasons they started using substances, difficulties or fears associated with leaving familial roles, stigma. They are five times more likely to have a history of sexual abuse.

### **3. MATERIALS AND METHOD**

#### **3.1 Research Objectives**

This study seeks to achieve the following four objectives:

RO1. Identify and analyze the temporal trends of DS applications in women's health research.

RO2. Map specific DS techniques/algorithms to women's health conditions addressed, evaluate the effectiveness, identify what works or does not, and determine alternative methods.

RO3. Examine the relationship between the scientific literature (SL) and usage and the citation impacts.

#### **3.2 Data Collection**

The data used in this study came from published documents selected through a literature survey performed on the scientific database index of the "ISI Web of Science (WoS) core collection." The search covers the period 2000 and 2020, which double as the effective trending era of DS and big data

analytics.<sup>11</sup> Our initial literature survey for this study occurred on February 22, 2021, with the final update on March 3, 2021. The complete search terms are as shown in Table 1. Limiting the search and selection of representative sample documents to the WoS was to ensure the quality of the publications included in this study and weeds out papers from unreliable sources.<sup>30</sup> This is particularly important for a trending topic like DS/big data and visual analytics that attracts a diverse research interest.<sup>11,31</sup> However, all the indexed articles on the WoS core collection addressing the topic of interest had an equal chance of being selected according to the preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.<sup>32</sup> Also, articles indexed on the WoS make up about 84% of the scholarship indexed in MEDLINE and SCOPUS.<sup>31,33</sup>

Table 1 Articles' Search, Filtering, Screening and Selection Criteria of Articles

Activities/Focus	Criteria
Search Term	TOPIC: (((("women") OR ("woman") OR ("gender")) AND ("health*") AND ("data science") OR ("machine learning") OR ("Analytics") OR ("visualization") OR ("algorithm")))). The search generated 3673 published documents.
Database Source	Web of Science Core Collection.
Years of Publication	Period covered/searched: 2000-2020. Articles dated 2021 were early access publications that occurred in 2020.
Publication Sources	All indexed sources in the Web of Science core collection that publish documents matching the search term.
Screening	None-medical related problems, studies on the general population; 3673 less 3110 = 563; non-data science and big data analytics methods 563 less 363 = 200 Articles addressing general medical issues from 200 less 50 = 150
Selection	All the unique documents filtered from the search and addressing the subject are included according to the criteria specified by the PRISMA statement. <sup>32</sup>

### 3.3 Filtering, Screening and Selection of the Scientific Literature

The unique 3,673 published articles retrieved from the Web of Science/Knowledge went through a thorough filtering and screening process based on the criteria in Table 1. Finally, the study included 150 screened published documents. The next steps involved extracting the relevant information in textual

form and exported to Excel and an R-studio programming environment for science mapping, quantitative analysis, and descriptive narratives.

### **3.4 Science Mapping and Visualization of Scientific Literature Using R-Base Open-Source Software**

The analysis involved different mapping of DS techniques to diverse women's health problems employs big data analytics and result visualization.<sup>42,108</sup> The study also analyzed the authors, countries, and regions where the study originates and the researchers' international collaborations. This helps to establish global awareness and utilization of modern DS techniques to improve women's health worldwide. We utilize an open-source bibliometric analysis package (Bibliometrix) embedded in the R programming Studio to achieve these goals.<sup>34</sup> The data used for the analysis are text files extracted from the 150 as explained in the above section. Further details about the open-source Bibliometrix and Vosviewer packages utilized for visualization are available elsewhere.<sup>34,35</sup> The approach is suited to a new field or subject, such as contemporary DS and visual analytics.<sup>11</sup>

## **4. RESULTS**

The summary of the data extracted from the published documents (Table 2) were generated using the R-programming Studio as described in the earlier section. The 150 papers included in the study appear in 113 different sources, indicating a broad range of publication outlets in diverse fields of research and practice. Also, 83.3% are published in journals, while 12% appear in the conference proceedings, both making 95.3% of the publications. The document per author (0.15) and the collaboration index (6.70)

indicates a high collaboration among authors, which is expected in a multidisciplinary field (health informatics).

Table 2 Summary of the scientific literature production, authors, documents, and the citation impact

Description	Results	Description	Results
Data coverage in years: 2000 to 2020; Active Period: 2006 to 2020			
Documents	150	Sources	113
Author Appearances	1201	Total Citation	1078
Authors/Single-authored documents	4	Average citations per documents	7.19
Authors – all authors/co-authors	997	Documents per Author	0.15
Authors of multi-authored documents	993	Co-Authors per Documents	7.85
Author Keywords	493	Authors Per Document	6.52
Collaboration Index	6.7	Keywords Plus (ID)	544
Types of Documents:			
Articles (Journals) 83.3%	125	Papers in Conference Proceedings (12%)	18
Letters (1.3%)	2	Meeting Abstract (3.3%)	5

#### 4.1 Temporal trends of Data Science (DS) Methods Application in Women’s Health Research

The analysis in this Section intends to achieve the first research objective (RO1) as defined in the previous section.

##### 4.1.1 Annual Scientific Literature Production (SLP)

The SLP measures the development of a research field.<sup>36</sup> The period 2000 to 2020 doubles as the emergence and growth of DS research and practice. The first article applying DS to solve Women’s health problems did not occur until 2006. The period 2006 to 2016 records just seven papers, while 143 SLP occurred from 2017-2020, indicating recent heightened interest (Figure 1). Studies using DS

methods to solve women’s health make 26.6% during the period, while 73.4% relate to general health issues or problems relating to men (Table 1; Figure 1). The low document per author (0.15) – see Table 2, indicates fewer studies using DS techniques to address women’s health problems. Figure 1 shows a complete trend of the SLP during the period.

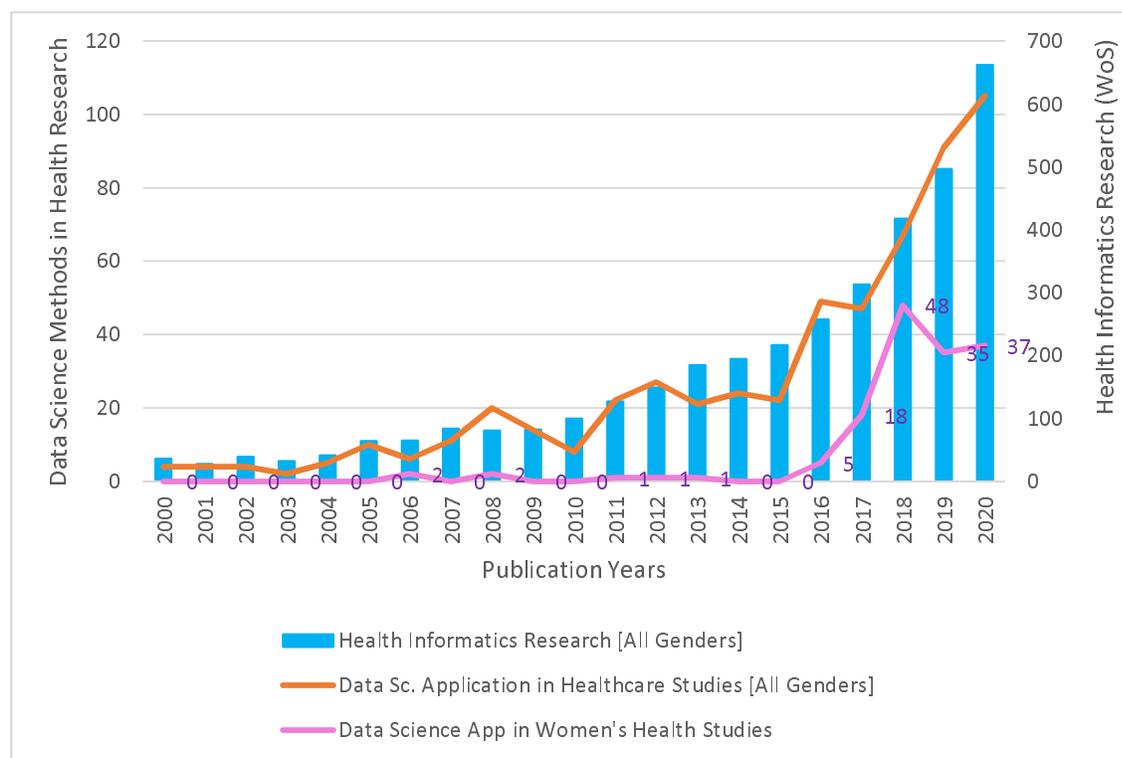


Figure 1: Annual scientific literature production on data science application in woman’s health

#### 4.1.2 Collaboration Network and Contributions of Countries and Institutions in DS in Women’s Health Studies

It is essential to know the institutions and countries pioneering the studies addressed in this paper as a pointer to the current spread and promoting more involvement. The 997 authors/co-authors come from 55 countries, with most of the authors domiciled in the USA (Table 2). Over 43% (65 out 150) of the SLP originate from research institutions in the USA, followed by India and Canada (Figure 2), both as

corresponding authors within the USA (MCP) and co-authors with outside country collaborators (SCP)<sup>11</sup>.

Figure 2 presents the top 20 countries.

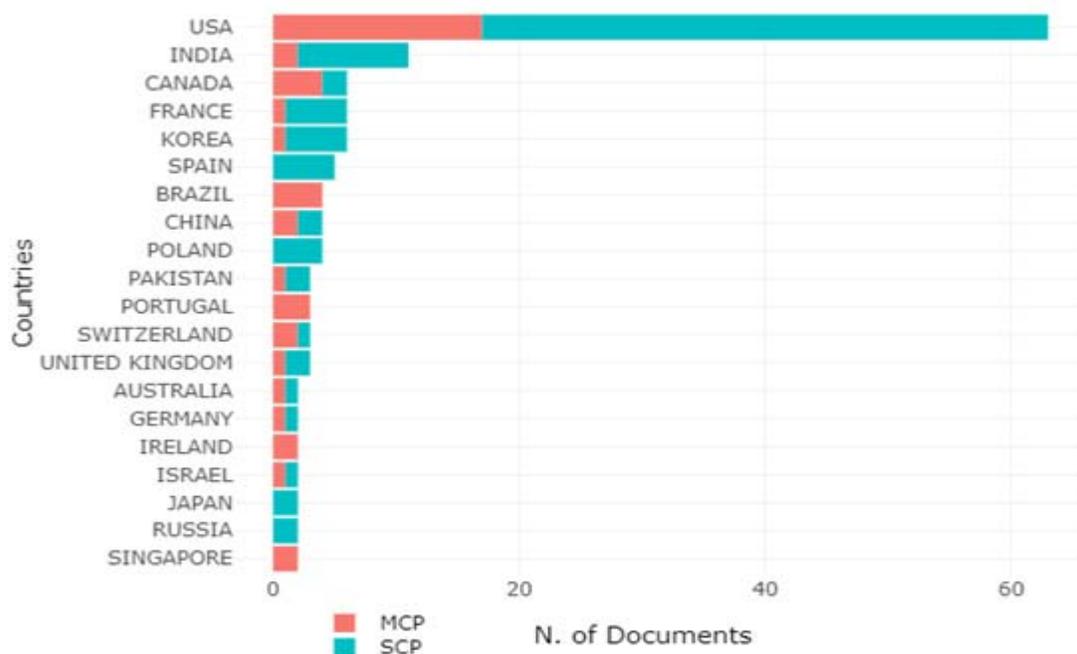


Figure 2. Countries Contributions of scientific literature and Collaborations (SCP: Corresponding Authors within the Country; MCP: Corresponding Authors Outside Country)

Regarding the contributing institutions, the most relevant and productive organizations originate from the USA (Figure 3), with the University of Washington, University of Pennsylvania, and Harvard Medical School, as the top three contributors. The nodes and the lines represent the contributions and connections, respectively.<sup>59</sup> Figure 3 presents the complete lists of institutions from different countries.

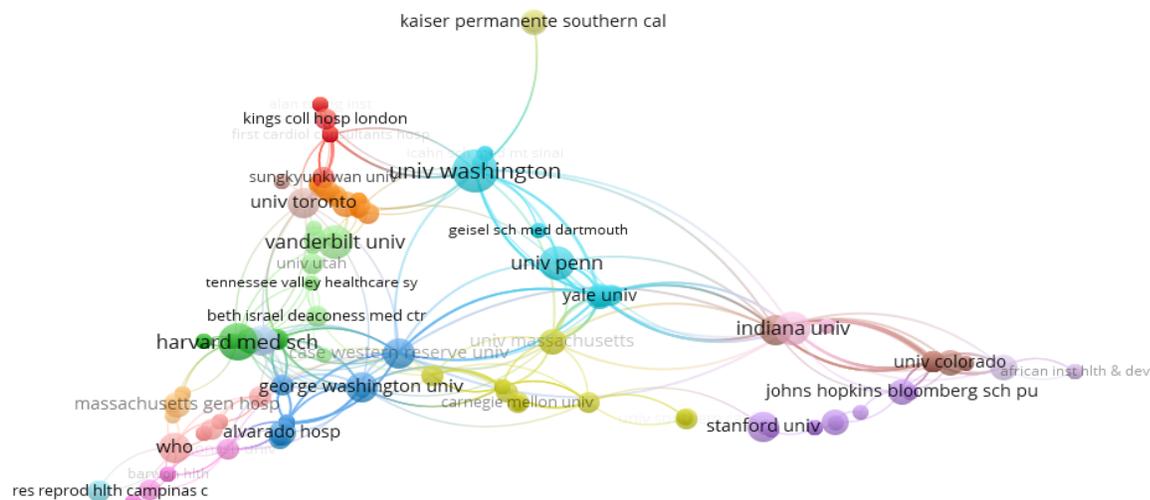


Figure 3. Institutions/Organizations involved in producing the scientific literature on data science application in women's health conditions.

#### 4.2 Science Mapping and Evaluation of Data Science Methods in Solving Women's Health Problems

The analysis in this Section intends to achieve the first research objective (RO2).

##### 4.2.1 Science Mapping of Data Science Methods and Women's Health Research, Authors, and Sources

Using an open-source bibliometric application embedded in R-Studio, we identify the top twenty-five (25) most frequent labels based on "authors' keywords" as the unit of analysis.<sup>33,59</sup> The authors' keywords make up the central focus of any scientific literature (SL), as demonstrated by studies elsewhere.<sup>31,45,59</sup> The result (Figure 4) shows machine learning as the most popular DS method used, while breast cancer, pregnancy issues, and other conditions, fall under gynecologic oncology.

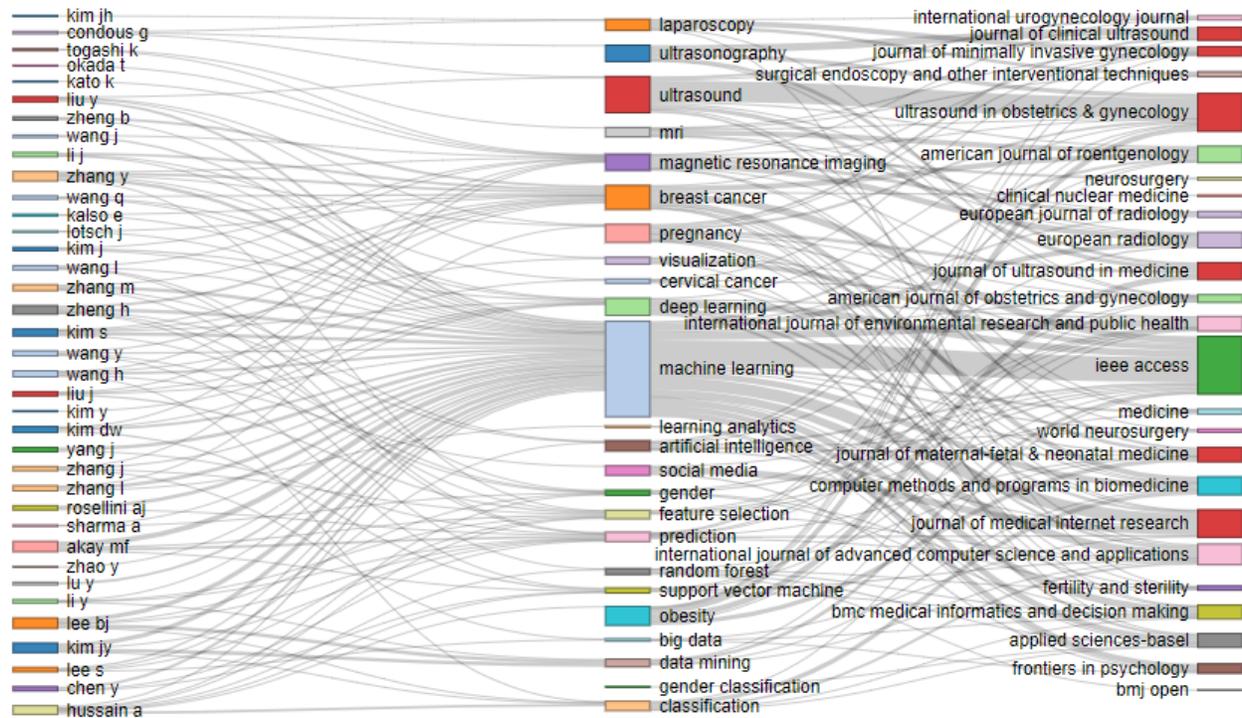


Figure 4. Top 25 data science methods and women's health problems mapped to the authors and sources.

The efforts indicate a strategic approach in using DS to address health conditions that place the most health burden on women.<sup>8,9</sup> On the sources, IEEE Access, Ultrasound in Obstetrics & Gynecology, JMIR publish most of the SLP on these subjects.

#### 4.2.2 Network Analysis and Visualization of DS Methods in Women's Health Studies

The results of network analysis using VOSVIWER, an open-source science mapping application, demonstrate the connections between the DS techniques, the diseases examined, and the inter-relationships among infections and illnesses. For example, "machine learning" is commonly used to study breast cancer and pregnancy-related problems. Similarly, the ongoing SARS-CoV-2 condition indicates a connection with COVID-19 and cardiovascular diseases (Figure 5). The science mapping utilizes the 493 synthesized authors' keywords (Table 2) as the unit of analysis and presents the



learning and novel algorithms, making it the most popularly used data science technique. Table 3 presents the complete results and descriptive analysis of the DS method and the problems tackled. These results echo the automatic mapping of the authors, keywords, and the sources (Figure 4/5) and network analysis that highlights the connections among the themes and topics using keywords as the unit of analysis (Figure 5).

Table 3. Data Science methods, problems tackled, and the application domain.

Women's Health Problems Tackled	Data Science and Analytics Techniques							
	AI	ML	DL	BDA	PA	NA	CDSS	VIS
<i>Breast Cancer</i>								
Predict and detect early breast cancer with accuracy and efficacy; identified false positive/negative.			[43]	[46] [123]				
Early detection of abnormalities in breasts, exam, and diagnosis in rural developing countries.		[37]						
Dec-support to clinicians (diagnosis & therapeutics).							[51]	[51]
Accurate detection/classification of malignant cells in breast cytology images (98.0% accurate/ effective).	[50]	[50]		[46]				
Preserve privacy of women with breast cancer.						[39]		
Estimate utility weights between elderly women with early-stage hormone receptor positive breast cancer.				[48]				
Early diagnosis: malignant neoplasms; visualization revealed 63% of breast cancer (80% at stages i-ii).								[42]
Predict long-term effect of oestrogen use on breast cancer & mortality (7645 postmenopausal women).			[49]					
Improves breasts cancer development risk estimation		[54]						
Predict 5-year breast cancer risk for prevention.		[52]			[52]			
Quick turnaround time for immunohistochemistry test and precise care of 402 breast cancer patients.				[40]				
Created care procedure to treat genitourinary syndrome of menopause for breast cancer survivors.						[41]		
Assign event date for a second breast cancer.						[118]		
<i>Gynecologic Oncology</i>								
Accurate identification of causative agents of vulvo-vaginitis to guide appropriate therapy.						[75]		[75]
Identifying uterine perforations and intrauterine device (IUD) expulsions and ascertaining breast feeding status during IUD insertion, utilizing sample of 282,028 women with IUD insertions.						[78]		
Determine that lactobacillus species found in African female like women of white decent; lactobacillus species found in the vagina of premenopausal healthy black women against inaccurate speculation						[76] [77]		

that they are susceptible to STDs and bacterial vaginosis due to lack of lactobacilli in their vagina.								
Identify/visualize the prevalence and risk factors of clitoral adhesions of 614 women.								[16]
Characterized cervical cancer burden and identified opportunities for prevention, and ovarian cancer.				[9]		[121]		
Detection and classification of cervical cells.			120]					
<i>Other Cancerous Condition</i>								
Predicts colorectal cancer among multi-ethnic colorectal cancer families (63% women); validates modifiable factors for cancer prevention.		[53]						
<i>Pregnancy</i>								
Identify ectopic pregnancy cases; 92.3% effectiveness.						[38]		
Identify causative factors/predict postpartum depression; early detection/intervention decisions; economic burden.		[47, 56, 69]		[67]		[68]	[47]	
Identifies severe maternal morbidity/risk factors.		[57]				[105]		
Identify the requirements for real time remote health monitoring of pregnant women.				[58]		[58]		
Predict/visualize early warning for risky pregnancy, especially for rural women as quality services are deficient.		[106]						[61]
Determine pregnancy related issues and health status of a newborn from demographic information, lifestyle (smoking and drug use, and other attributes during gestation period); Accurately assess gestational age of babies compared to the early ultrasound date.		[114]				[111]		
Identify causes of stillbirth in low-income countries; a rate of 27.2 per 1000 births: Asphyxia: 46.6% (38% had prolonged or obstructed labor); infection (20.8%); congenital anomalies due to Syphilis; prematurity (6.6%); antepartum hemorrhage: 19%; pre-eclampsia /eclampsia; 67.4% of the stillbirths had no signs of maceration; obstetric hemorrhage, pregnancy-related infection. Pregnancy/postpartum increased the risk of venous thromboembolism; worry plays a unique role in pre-eclampsia; predict stillbirths among 4.9m data.		[113]				[70] [63] [71] [119] [127]		
Pregnant women/nCoV infection: see COVID section.		[102]				[100]		
Identify women with symptomatic uterine fibroids.						[66]		
Better classification from buccal swabs of mothers from which DNA can be extracted that correctly classified mothers/child carrying mosaicism.		[107]						
Create digital tool for antenatal care: algorithms for developing clinical workflows and testing.				[125]				

Created a smart mobile data module for the remote monitoring and visualization of fetal heart rate.		[115]						[115]
Proposed methods of handling uncertainty in the decision-making related to high-risk pregnancy.								
Predicted preterm birth; miscarriage and results transmitted to doctors and pregnant woman for quick decision making.		[109]		[110]			[110]	
Determine the risks around the neighborhood prevalence of preterm birth mediated by infection and substance use.		[112]						
Monitoring and identifying high risk pregnancy / outcome.				[71]		[72]		
Pregnant women use of online forums to discuss maternal health, baby-related topics, people and relationships and pregnancy state pain.		[69]		[69]				
Predict/identify obesity among pregnant women, pre-pregnancy body mass index, child's sex, effects of smoking during pregnancy (subject: 2334 healthy non obese pregnant women).					[79]	[44]		
Identified pregnancy episodes, start dates, delivery outcomes with live birth for 4.9 million women; achieved high accuracy (92.8% - 96.4%); Predicts fetal congenital anomalies 89.5% accuracy.		[62]				[64, 65]		
<i>Kidney Diseases</i>								
Determines the difference between re-calculated estimated glomerular filtration rate on retrospective data and the use of automated value. Kidney function assessment and staging of kidney diseases.				[124]				
<i>Alzheimer's Disease</i>								
Determine risk factors associated with Alzheimer's disease in 350 ageing women using unsupervised ML.		[80]						
<i>Cardiovascular Disease</i>								
Predicts 5-year risk of incident cardiovascular disease without the aid of clinical measures achieving successes within 95% confidence interval for women.						[81]		
Identify predictors of incident myocardial infarction in women's health initiative cohort of 46,568 patients.		[82]						
Determine whether the socio-economic gradient of estimated cardiovascular risk differs between ethnic groups (see socioeconomics).						[83]		
<i>Stress and Mental-Health</i>								
Physiological and behavioral measures captured through wearable sensors and mobile phones, for self-reported stress and mental health status of 201 college students (36% female).		[84]						
<i>Sexual Health</i>								

Diagnose/treatment hypoactive sexual desire disorder.						[85]		
Classified women into different health risk segments for customized awareness programs to resolve serious health issues, e.g., sexually transmitted infection.		[13]						
<i>Radiology</i>								
Measuring lung vessel morphology and visualization (55% of the sample were women).						[86]		[86]
<i>HIV/AIDS</i>								
Determine predictors of poor health outcomes among HIV-infected inpatients. Of 66% women, 53% had tuberculosis confirmed.						[55]		
<i>Urinary Tract Infection</i>								
Identifies women with urinary incontinence. Symptoms includes painful urges; pyelonephritis; three or more urinary tract infections in the last 12 months; dysuria; visible hematuria.						[87]		
<i>Menstrual Cycle</i>								
Food visualization elicited greater brain inactivation; intake of food during menstrual cycle fluctuates especially during early follicular and luteal phases.								[88]
Investigated joint model for serum hepcidin and iron during menstrual cycle.						[89]		
<i>Occupational Health</i>								
Unveil health risk factors in women at workplace and designing future medical protocols/policies.	[90]							
Arthritis associated physical stressors (heavy physical work & awkward trunk posture), kneeling and arm elevation (21,389 sample (49.2% women)).					[91]			
Identified factors with greatest impact on worker's health surveillance process (lung auscultation and breathing exploration); women show more resilience on occupational disease than men.		[14]						
<i>Socioeconomic Factors on Women's Health</i>								
Predicts social determinants of healthy aging and maternal healthcare utilization in 5565 women.		[104]			[92]			
See cardiovascular risk differences by ethnic groups.						[83]		
Pattern recognition of minority/women experienced more disability and mental illness.						[93]		
Identify severe maternal morbidity and comorbidities based on sociodemographic characteristics.						[105, 120]		
Accuracy of sex classification to enhance women's health/safety using gait information.	[94]					[94]		
<i>General Health Checks</i>								
Evaluates inter- and intra-rater measured reliability of cyclotorsion in healthy subjects (56% women)						[96]		

Visualizing electronic health records reveals sex and age biases in admin known drug-drug interactions.								[97]
Confirmed that RNA expression data is limited and cannot be used to identify specific patients.		[98]						
Predicts risk factors in stroke risk model (age, bmi, cholesterol, hypertension, diabetes, smoking status and intensity, alcohol drinking); 82% in women.					[99]			
Women’s worries: heart disease & breast cancer.	[113]					[117]		[95]
Social/behavioral determinants of health are helpful to understand health and health care outcomes.								[73]
Predict gestation diabetes in 4378 women (98.4% effective).						[101]		
Identify genes that discriminate individuals with fibromyalgia syndrome diagnosis from healthy controls in women.					[116]			
<i>Coronavirus Disease (COVID-19)</i>								
Pregnant women diagnosed of COVID-19 can have severe morbidities.						[100]		
Identified gaps in women v men; (51% of patients with diagnosed COVID-19 cases were women; headache, anosmia, and ageusia more frequent in females. Chest X-ray & blood tests were performed less frequently in women (65.5% vs. 78.3% and 49.5% vs. 63.7%))		[102]						
Sex gap among 583 healthcare workers (72% women). Women more vulnerable, being impacted by insomnia, depression, poor well-being		[126]						

#### 4.3 The Relationship Between the Scientific Literature (SL) and Usage and Citation Analysis

Here we examine the articles' usage and citation impacts. The data from the selected 150 indexed documents on the WoS come in two streams, namely, U1 and U2 (Table 4).

- U1 (usage count within the past 180 days or last six months): 53 articles used 145 times.
- U2 (usage count in the past 7 years, 2014-2020): 121 articles used 806 times.

The results (U1=35.3%, U1=81%), indicating active usage of the scientific literature during the period shown. The data does not distinguish specific users (researchers, health/industry practitioners, and

others) or for what purposes. The frequency of the articles' usage also increased astronomically in the most recent years (U1=98%, U1=91% used between 2017-2020), indicating an active research area.

Table 4. Usage and citation impact of the scientific literature on data science application in women's health studies (U1, U2 = usage in the last 6 months, 7 years, respectively).

REF	Article Focus	Sources	PY	Citation (WoS)	U1	U2
[49]	Breast cancer mortality	Lancet Oncology	2012	221	0	25
[76]	Lactobacillus vaginal microbiota	Sexually Transmitted Diseases	2006	62	0	3
[84]	Stress and mental health	J of medical Internet research	2018	47	3	37
[85]	Sexual desire disorder in women	Mayo Clinic Proceedings	2018	36	2	14
[88]	Brain responses to food images during menstrual cycle in healthy young women	American Journal of Clinical Nutrition	2011	36	0	12
[41]	Genitourinary syndrome of menopause in women with high risk for breast cancer	Menopause	2018	36	1	8
[100]	Care of pregnant women with COVID-19 in labor & delivery	American Journal of Obstetrics and Gynecology	2020	27	6	8
[119]	Syphilis during pregnancy and threat to maternal-fetal health	American journal of obstetrics and gynecology	2017	25	0	17
[104]	Factors determining healthy ageing	Scientific Reports	2017	23	0	6
[43]	Predicting breast cancer	Radiology	2019	22	3	21
[105]	Severe maternal morbidity and health disparities	Obstetrics and gynecology	2017	22	0	7
[70]	Stillbirth in low-and middle-income countries	International Journal of Obstetrics & Gynecology	2018	22	0	5
[79]	Organic pollutants and gestational diabetes among healthy US women.	Environment international	2019	18	4	22
[94]	Accurate sex classification to enhance women's health/safety.	Expert Systems with Applications	2018	15	0	42
[120]	Detection and classification of cervical cells	Journal of Supercomputing	2020	15	18	18
[81]	Predicts risk of cardiovascular disease for women.	CMAJ	2018	13	0	3
[50]	Detection and classification of breast cancer	Future Generation Computer Systems	2019	11	0	5
[63]	Maternal deaths in low / middle-income countries	International Journal of Obstetrics & Gynaecology	2018	11	1	5
[123]	Overcoming false positives and false diagnoses of breasts tumor	Journal of Grid Computing	2018	10	0	18
[101]	Gestational diabetes mellitus prediction in early pregnancy.	Scientific reports	2019	10	0	7

Regarding the citation impact, the total basic citation count was 1078, an average of 7.18 per document. However, 49 papers (32.7%) are yet to receive any citations, while 68.3% gained at least one citation, ranging as follows (1-10: 53.3%; 11-65: 14%; >65: <1% citations). Table 4 presents the top 20 most cited documents and the focus of women's health research.

## 5. DISCUSSION

This study evaluates the DS methods employed in solving diverse women's problems, including machine learning, big data analytics, deep learning, and novel/existing algorithms. Other techniques used are clinical decision support systems and visual analytics/visualization. These techniques were very effective in tackling diverse women's health problems. Significant and improved health outcomes occurred in predicting and managing breast cancer, gynecologic oncology, and pregnancy-related health challenges (pre-eclampsia, eclampsia, prediction of risks during pregnancy, and more). Other health challenges addressed include menstrual cycle, psychosocial stress, gestational diabetes, and vaginosis. Diseases that are not solely related to women, like Alzheimer's, cardiovascular diseases, and even scabies, are also examined. Occupational health and the importance of women's health records confidentiality, and utilization of health facilities, family planning, and the characterization of drug-drug interaction. Table 3 presents a detailed analysis.

The research employing DS involves data collection through primary and secondary sources (medical records, claims data, and questionnaire administration) involving large volumes and sometimes complex data structures. The studies involved diverse groups of women, including patients, college students, healthcare and other workers, pregnant women, women living in a specified neighborhood, and patient caregivers. The sample sizes also vary for the different studies. The most extensive cohort study involved identifying pregnancy episodes and outcomes using a claims database of about 4.9

million women and the smallest that investigated brain activation during food visualization and the changes in estradiol concentration among nine women.<sup>88,127</sup>

A significant breakthrough in women's health using DS occurred in using big data and visual analytics to undertake studies in obstetrics and gynecology, which helps to reduce maternal mortality.<sup>11,128</sup> The methods use personal information such as sex and genomic data to predict cancer.<sup>98</sup> However, there are ongoing challenges and societal hurdles, as some women seem uncomfortable using big data to develop health risk prediction.<sup>129</sup> We believe that formulating appropriate privacy policies using big secondary data can resolve these issues.

The study also identifies disparities between women and minority groups. For example, a sex gap was identified in the treatment of COVID-19 among hospital patients. Although 51% of patients with diagnosed COVID-19 cases were women, more men were treated than women. Chest X-ray and blood tests were performed less frequently in women (65.5% vs. 78.3% and 49.5% vs. 63.7%).<sup>102</sup> Issues relating to access to health are common, including women of different groups. Healthcare utilization differed across groups of cervical cancer patients, and this revelation can help identify opportunities for early prevention.<sup>9</sup> A study found wealth, social group, literacy, religion, and early age at marriage as social determinants of maternal health care utilization.<sup>92</sup> Midlife women's health and analysis of most women's concerns involves worries about breast cancer and heart disease.<sup>95</sup>

## 6. CONCLUSION

This study focused on evaluating DS methods and algorithms to solve women's health problems. The goals in evaluating the techniques used, maps the methods and issues, analyzing the effectiveness, identifying what works, and what else can help to mitigate the disparities regarding women's health burden and access to health. This paper has achieved these goals, further determines the origin and growth of this area of study, and predicts future growth using simple linear trends and forecasting.

The science mapping of the scientific literature production on this subject indicates an upward trend. About 95% (143 out of 150) of research utilizing data science methods in women's health studies occurred within the last four (4) years (2017 to 2020), while only 4.7% of the research happened between 2000 and 2016, indicating a new field. The potency of the data science methods and the improved outcomes that can help solve women's health problems raises hope of reducing the health gap between men and women, especially regarding the disease burden. However, addressing access to health requires more than advances in computer technology and the deployment of information systems and methods. Policymakers need to play a critical role in reducing the overall gaps, especially access to health. Further, the ratio of the studies involving data science to solving women's health issues is relatively low and requires greater participation. Ironically, while the data science techniques appear promising to address women's health problems, but some women activists are unknowingly opposing the adoption, indicating the need for enlightenment regarding the usefulness of the identified techniques.

## References

1. Bush M, Lederer AL, Li X, Palmisano J, Rao S. The alignment of information systems with organizational objectives and strategies in health care. *International journal of medical informatics*. 2009 Jul 1;78(7):446-56.
2. Capan M, Khojandi A, Denton BT, Williams KD, Ayer T, Chhatwal J, Kurt M, Lobo JM, Roberts MS, Zaric G, Zhang S. From data to improved decisions: operations research in healthcare delivery. *Medical Decision Making*. 2017 Nov;37(8):849-59.
3. Mutlag AA, Abd Ghani MK, Arunkumar NA, Mohammed MA, Mohd O. Enabling technologies for fog computing in healthcare IoT systems. *Future Generation Computer Systems*. 2019 Jan 1;90:62-78.
4. Akpan IJ. An empirical study of the impacts of virtual reality on discrete-event simulation (Doctoral dissertation), University of Lancaster. <https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.440409>
5. Günal MM, Pidd M. Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*. 2010 Mar;4(1):42-51.
6. Cook NL. Eliminating the sex and gender gap and transforming the cardiovascular health of all women. *Ethnicity & disease*. 2019;29(Suppl 1):65.

7. Wisdom JP, McGee MG, Horner-Johnson W, Michael YL, Adams E, Berlin M. Health disparities between women with and without disabilities: a review of the research. *Social work in public health*. 2010 Apr 28;25(3-4):368-86.
8. Sina BJ. Pregnancy and the global disease burden. *Reproductive Health*. 2017 Dec;14(3):5-8.
9. Pruitt SL, Werner CL, Borton EK, Sanders JM, Balasubramanian BA, Barnes A, Betts AC, Skinner CS, Tiro JA. Cervical cancer burden and opportunities for prevention in a safety-net healthcare system. *Cancer Epidemiology and Prevention Biomarkers*. 2018 Dec 1;27(12):1398-406.
10. Haux R. Medical informatics: past, present, future. *International journal of medical informatics*, 2010; 79(9), 599-610.
11. Akpan IJ, Akpan AA. Multiple criteria analysis of the popularity and growth of research and practice of visual analytics, and a forecast of the future trajectory. *International Transactions in Operational Research*. 2021; 1-24. DOI:10.1111/itor.12952.
12. Akoglu L, Tong H, Koutra D. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery*. 2015 May;29(3):626-88.
13. Anbu S, Sarmah B. Machine learning approach for predicting womens health risk. In 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS) 2017 Jan 6 (pp. 1-4). IEEE.
14. Abad A, Gerassis S, Saavedra Á, Giráldez E, García JF, Taboada J. A Bayesian assessment of occupational health surveillance in workers exposed to silica in the energy and construction industry. *Environmental Science and Pollution Research*. 2019 Oct;26(29):29560-9.
15. Akpan IJ, Shanker M. The confirmed realities and myths about the benefits and costs of 3D visualization and virtual reality in discrete event modeling and simulation: A descriptive meta-analysis of evidence from research and practice. *Computers & Industrial Engineering*. 2017 Oct 1;112:197-211.
16. Aerts L, Rubin RS, Randazzo M, Goldstein SW, Goldstein I. Retrospective study of the prevalence and risk factors of clitoral adhesions: women's health providers should routinely examine the glans clitoris. *Sexual medicine*. 2018 Jun 1;6(2):115-22.
17. deKleijn M, Lagro-Janssen AL, Canelo I, Yano EM. Creating a roadmap for delivering gender-sensitive comprehensive care for women Veterans: results of a national expert panel. *Medical Care*. 2015 Apr;53(4 Suppl 1): S156.
18. Keen MU, Reddivari AK. Osteoporosis In Females. *StatPearls [Internet]*. 2020 Jun 23.
19. van Vollenhoven, R.F., 2009. Sex differences in rheumatoid arthritis: more than meets the eye... *BMC medicine*, 7(1), pp.1-4. DOI:10.1186/1741-7015-7-12
20. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Das SR, Delling FN. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation*. 2019; 139(10): e56-28.
21. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. *Journal of the American College of Cardiology*, 1993; 22(4S1), A6-A13.
22. Kim HJ, Lim H, Moon A. Sex differences in cancer: epidemiology, genetics and therapy. *Biomolecules & therapeutics*. 2018 Jul;26(4):335.
23. Kessler RC, Crum RM, Warner LA, Nelson CB, Schulenberg J, Anthony JC. Lifetime co-occurrence of DSM-III-R alcohol abuse and dependence with other psychiatric disorders in the National Comorbidity Survey. *Archives of general psychiatry*. 1997; 54(4):313-21.
24. Conway KP, Compton W, Stinson FS, Grant BF. Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry*. 2006 Feb 15;67(2):247-57.
25. Green CA. Gender and use of substance abuse treatment services. *Alcohol Research & Health*. 2006; 29(1):55.
26. Green T, Flash S, Reiss AL. Sex differences in psychiatric disorders: what we can learn from sex chromosome aneuploidies. *Neuropsychopharmacology*. 2019; 44(1):9-21.
27. Cameron KA, Song J, Manheim LM, Dunlop DD. Gender disparities in health and healthcare use among older adults. *Journal of women's health*. 2010 Sep 1;19(9):1643-50.
28. Weissman J, Russell D, Jay M, Malaspina D. Racial, ethnic, and gender disparities in health care access and use among US adults with serious psychological distress. *Psychiatric Services*. 2018 May 1;69(5):517-22.

29. Ferraz-Torres M, Belzunegui-Otano T, Marín-Fernandez B, Martínez-García Ó, Ibañez-Beroiz B. Differences in the treatment and evolution of acute coronary syndromes according to gender: what are the causes? *Journal of clinical nursing*. 2015 Sep;24(17-18):2468-77.
30. Jacso P. As we may search—comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current science*, 2008; 89(9), 1537-1547.
31. Akpan IJ, Aguolu OG, Ezeume IC. Overcoming the challenge of communicating the concept and science of SARS-CoV-2 and COVID-19 to non-experts. *Challenge*, 2021; 1-24. <https://doi.org/10.1080/05775132.2021.1912984>.
32. Moher M, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151(4): 264–269.
33. Akpan IJ. Scientometric evaluation and visual analytics of the scientific literature production on entrepreneurship, small business ventures, and innovation. *Journal of Small Business & Entrepreneurship*. 2020 Jul 8:1-29. DOI: 10.1080/08276331.2020.1786229.
34. Aria M, Cuccurullo C (2016) Bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr* 11:959–975.
35. Van Eck, N. J., & Waltman, L. (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping." *Scientometrics*, 84 (2): 523–538. doi:10.1007/s11192-009-0146-3.
36. Larivière V, Archambault É, Gingras Y. Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and technology*. 2008 Jan 15;59(2):288-96.
37. Pinder LF, Nzayisenga JB, Shibemba A, Kusweje V, Chiboola H, Amuyunzu-Nyamongo M, Kapambwe S, Mwaba C, Lermontov P, Mumba C, Henry-Tillman R. Demonstration of an algorithm to overcome health system-related barriers to timely diagnosis of breast diseases in rural Zambia. *PLoS One*. 2018 May 10;13(5):e0196985.
38. Getahun D, Shi JM, Chandra M, Fassett MJ, Alexeeff S, Im T. M, ... & Raine-Bennett T. Identifying Ectopic Pregnancy in a Large Integrated Health Care Delivery System: Algorithm Validation. *JMIR medical informatics*, 2020; 8(11), e18559.
39. Vadavalli A, Subhashini R. An Improved Differential Privacy-Preserving Truth Discovery approach In Healthcare. In 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) 2019 Oct 17 (pp. 1031-1037). IEEE.
40. López-Pineda A, Rodríguez-Moran MF, Álvarez-Aguilar C, Valle SM, Acosta-Rosales R, Bhatt AS, Sheth SN, Bustamante CD. Data mining of digitized health records in a resource-constrained setting reveals that timely immunophenotyping is associated with improved breast cancer outcomes. *BMC cancer*. 2018 Dec;18(1):1-0.
41. Faubion SS, Larkin LC, Stuenkel CA, Bachmann GA, Chism LA, Kagan R, Kaunitz AM, Krychman ML, Parish SJ, Partridge AH, Pinkerton JV. Management of genitourinary syndrome of menopause in women with or at high risk for breast cancer: consensus recommendations from The North American Menopause Society and The International Society for the Study of Women's Sexual Health. *Menopause*. 2018 Jun 1;25(6):596-608.
42. Gamirov R, Komarova L, Khasanov R, Gamirova R. Early Diagnosis of Breast Cancer: Public Health and Mammographic Screening. *BioNanoScience*. 2018 Mar;8(1):407-11.
43. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, Barkan E, Herzel E, Naor S, Karavani E, Koren G. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*. 2019 Aug;292(2):331-42.
44. Saijo Y, Ito Y, Yoshioka E, Sato Y, Minatoya M, Araki A, Miyashita C, Kishi R. Identifying a risk score for childhood obesity based on predictors identified in pregnant women and 1-year-old infants: An analysis of the data of the Hokkaido Study on Environment and Children's Health. *Clinical Pediatric Endocrinology*. 2019;28(3):81-9.
45. Akpan IJ, Ezeume IC. The Challenges Faced by Parents and Children from Divorce. *Challenge*. 2020 Nov 1; 63(6):365-77. DOI:10.1080/05775132.2020.1842025.
46. Gomathi N, Srinivasulu P, Nikhil CV, Vineeth KV. Interactive Dashboard for The Betterment of Patient Health Using Big Data Analytics. *RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES*. 2016 Nov 1;7(6):1697-707.
47. Zhang Y, Joly R, Hermann A, Pathak J. Using Machine Learning and Electronic Health Records to Predict Postpartum Depression [38E]. *Obstetrics & Gynecology*, 2020; 135, 59S-60S.

48. Ali AA, Xiao H, Tawk R, Campbell E, Semykina A, Montero AJ, Diaby V. Comparison of health utility weights among elderly patients receiving breast-conserving surgery plus hormonal therapy with or without radiotherapy. *Current medical research and opinion*. 2017 Feb 1;33(2):391-400.
49. Anderson GL, Chlebowski RT, Aragaki AK, Kuller LH, Manson JE, Gass M, Bluhm E, Connelly S, Hubbell FA, Lane D, Martin L. Conjugated equine oestrogen and breast cancer incidence and mortality in postmenopausal women with hysterectomy: extended follow-up of the Women's Health Initiative randomised placebo-controlled trial. *The lancet oncology*. 2012 May 1;13(5):476-86.
50. Khan SU, Islam N, Jan Z, Din IU, Khan A, Faheem Y. An e-Health care services framework for the detection and classification of breast cancer in breast cytology images as an IoMT application. *Future Generation Computer Systems*. 2019 Sep 1;98:286-96.
51. Sekar B, Lamy JB, Muro N, Pinedo AU, Seroussi B, Larburu N, Guézennec G, Bouaud J, Masero FG, Arrúe M, Wang H. Intelligent clinical decision support systems for patient-centered healthcare in breast cancer oncology. In 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom) 2018 Sep 17 (pp. 1-6). IEEE.
52. Stark GF, Hart GR, Nartowt BJ, Deng J. Predicting breast cancer risk using personal health data and machine learning models. *Plos one*. 2019 Dec 27;14(12):e0226765.
53. Shiao SP, Grayson J, Lie A, Yu CH. Predictors of the healthy eating index and glycemic index in multi-ethnic colorectal cancer families. *Nutrients*. 2018 Jun;10(6):674.
54. Pérez-Benito FJ, Signol F, Perez-Cortes JC, Pollán M, Pérez-Gómez B, Salas-Trejo D, Casals M, Martinez I, Llobet R. Global parenchymal texture features based on histograms of oriented gradients improve cancer development risk estimation from healthy breasts. *Computer methods and programs in biomedicine*. 2019 Aug 1;177:123-32.
55. Griesel R, Stewart A, van der Plas H, Sikhondze W, Mendelson M, Maartens G. Prognostic indicators in the World Health Organization's algorithm for seriously ill HIV-infected inpatients with suspected tuberculosis. *AIDS research and therapy*. 2018 Dec;15(1):1-9.
56. Wang S, Pathak J, Zhang Y. Using electronic health records and machine learning to predict postpartum depression. *Studies in health technology and informatics*. 2019;264:888-92.
57. Gao C, Osmundson S, Yan X, Edwards DV, et al. Learning to identify severe maternal morbidity from electronic health records. *Studies in health technology and informatics*. 2019 21;264:143.
58. Kumar S, Gupta Y, Mago V. Health-monitoring of pregnant women: Design requirements, and proposed reference architecture. In 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC) 2019 Jan 11 (pp. 1-6). IEEE.
59. Yan E, Ding Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and cword networks relate to each other. *Journal of the American Society for Information Science and Technology*. 2012 63(7):1313-26.
60. Akpan IJ, Brooks RJ. Experimental evaluation of user performance on two-dimensional and three-dimensional perspective displays in discrete-event simulation. *Decision Support Systems*. 2014 Aug 1; 64:14-30. <https://doi.org/10.1016/j.dss.2014.04.002>.
61. Dijaya R, Cholifah C, Djauharoh D, Nisak U, Syahminan S. Visual cohort baby recording based on internet of things for maternal and child health service. In *Journal of Physics: Conference Series* 2019 Dec 1 (Vol. 1402, No. 7, p. 077038). IOP Publishing.
62. Akbulut A, Ertugrul E, Topcu V. Fetal health status prediction based on maternal clinical history using machine learning techniques. *Computer methods and programs in biomedicine*. 2018 1;163:87-100.
63. Pasha O, McClure EM, Saleem S, Tikmani SS, Lokangaka A, Tshetu A, Bose CL, Bauserman M, Mwenechanya M, Chomba E, Carlo WA. A prospective cause of death classification system for maternal deaths in low and middle-income countries: results from the Global Network Maternal Newborn Health Registry. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2018 Aug;125(9):1137-43.
64. Blotière PO, Weill A, Dalichampt M, Billionnet C, Mezzarobba M, Raguideau F, Dray-Spira R, Zureik M, Coste J, Alla F. Development of an algorithm to identify pregnancy episodes and related outcomes in health care claims databases: an application to antiepileptic drug use in 4.9 million pregnant women in France. *Pharmacoepidemiology and drug safety*. 2018 Jul;27(7):763-70.

65. Ishikawa T, Obara T, Nishigori H, Miyakoda K, Inoue R, Hoshiai T, Saito M, Yaegashi N, Mano N. Development of algorithms to determine the onset of pregnancy and delivery date using health care administrative data in a university hospital in Japan. *Pharmacoepidemiology and drug safety*. 2018 Jul;27(7):751-62.
66. Hoffman SR, Vines AI, Halladay JR, Pfaff E, Schiff L, Westreich D, Sundaresan A, Johnson LS, Nicholson WK. Optimizing research in symptomatic uterine fibroids with development of a computable phenotype for use with electronic health records. *American journal of obstetrics and gynecology*. 2018 Jun 1;218(6):610-e1.
67. Taylor CL, Munk-Olsen T, Howard LM, Vigod SN. Schizophrenia around the time of pregnancy: leveraging population-based health data and electronic health record data to fill knowledge gaps. *BJPsych open*. 2020 Sep;6(5).
68. Epperson CN, Huang MY, Cook K, Gupta D, Chawla A, Greenberg PE, Eldar-Lissai A. Healthcare resource utilization and costs associated with postpartum depression among commercially insured households. *Current Medical Research and Opinion*. 2020 Oct 2;36(10):1707-16.
69. Wexler A, Davoudi A, Weissenbacher D, Choi R, O'Connor K, Cummings H, Gonzalez-Hernandez G. Pregnancy and health in the age of the internet: a content analysis of online "birth club" forums. *PLoS one*. 2020 Apr 14;15(4):e0230947.
70. McClure EM, Garces A, Saleem S, Moore JL, Bose CL, Esamai F, Goudar SS, Chomba E, Mwenechanya M, Pasha O, Tshefu A. Global Network for Women's and Children's Health Research: probable causes of stillbirth in low-and middle-income countries using a prospectively defined classification system. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2018 Jan;125(2):131-8.
71. Moreira MW, Rodrigues JJ, Carvalho FH, Chilamkurti N, Al-Muhtadi J, Denisov V. Biomedical data analytics in mobile-health environments for high-risk pregnancy outcome prediction. *Journal of Ambient Intelligence and Humanized Computing*. 2019 Oct;10(10):4121-34.
72. Moreira MW, Rodrigues JJ, Kumar N, Al-Muhtadi J, Korotaev V. Nature-inspired algorithm for training multilayer perceptron networks in e-health environments for high-risk pregnancy care. *Journal of medical systems*. 2018 Mar;42(3):1-0.
73. Monsen KA, Brandt JK, Brueshoff BL, Chi CL, Mathiason MA, Swenson SM, Thorson DR. Social determinants and health disparities associated with outcomes of women of childbearing age who receive public health nurse home visiting services. *Journal of Obstetric, Gynecologic & Neonatal Nursing*. 2017 Mar 1;46(2):292-303.
74. Akpan, I. J., D. Soopramanien, and D. H. Kwak. 2020. "Cutting-Edge Technologies for Small Business and Innovation in the Era of COVID-19 Global Health Pandemic." *Journal of Small Business & Entrepreneurship*: 1–11. doi:10.1080/08276331.2020.1799294.
75. Ackerman SJ, Knight T, Wahl PM, Cartwright CP. Health care utilization and costs following amplified versus non-amplified molecular probe testing for symptomatic patients with suspected vulvovaginitis: a US commercial payer population. *ClinicoEconomics and outcomes research: CEOR*. 2019;11:179.
76. Anukam KC, Osazuwa EO, Ahonkhai I, Reid G. Assessment of Lactobacillus species colonizing the vagina of apparently healthy Nigerian women, using PCR-DGGE and 16S rRNA gene sequencing. *World Journal of Microbiology and Biotechnology*. 2006 Oct;22(10):1055-60.
77. Anukam KC, Osazuwa EO, Ahonkhai I, Reid G. Lactobacillus vaginal microbiota of women attending a reproductive health care service in Benin city, Nigeria. *Sexually transmitted diseases*. 2006;33(1):59-62.
78. Anthony MS, Armstrong MA, Getahun D, Scholes D, Gatz J, Schulze-Rath R, Postlethwaite D, Merchant M, Alabaster AL, Chillemi G, Raine-Bennett T. Identification and validation of uterine perforation, intrauterine device expulsion, and breastfeeding in four health care systems with electronic health records. *Clinical epidemiology*. 2019;11:635.
79. Rahman ML, Zhang C, Smarr MM, Lee S, Honda M, Kannan K, Tekola-Ayele F, Louis GM. Persistent organic pollutants and gestational diabetes: a multi-center prospective cohort study of healthy US women. *Environment international*. 2019 Mar 1;124:249-58.
80. Riedel BC, Daianu M, Ver Steeg G, Mezher A, Salminen LE, Galstyan A, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. Uncovering biologically coherent peripheral signatures of health and risk for Alzheimer's disease in the aging brain. *Frontiers in aging neuroscience*. 2018 Nov 29;10:390.
81. Manuel DG, Tuna M, Bennett C, Hennessy D, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). *Cmaj*. 2018 Jul 23;190(29):E871-82.

82. Avram R, Tison G, Nah G, Howard BV, et al. Using Machine Learning Methods to Identify Predictors of Incident Myocardial Infarction in the Women's Health Initiative Cohort. *Circulation*. 2018 6;138(Suppl\_1): A15034-.
83. Perini W, Agyemang C, Snijder MB, Peters RJ, Kunst AE. Ethnic disparities in educational and occupational gradients of estimated cardiovascular disease risk: The Healthy Life in an Urban Setting study. *Scandinavian journal of public health*. 2018 Mar;46(2):204-13.
84. Sano A, Taylor S, McHill AW, Phillips AJ, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. *Journal of medical Internet research*. 2018;20(6):e210.
85. Clayton AH, Goldstein I, Kim NN, Althof SE, Faubion SS, Fought BM, Parish SJ, Simon JA, Vignozzi L, Christiansen K, Davis SR. The International Society for the Study of Women's Sexual Health process of care for management of hypoactive sexual desire disorder in women. In *Mayo Clinic Proceedings* 2018 Apr 1 (Vol. 93, No. 4, pp. 467-487). Elsevier.
86. Pienn M, Burgard C, Payer C, Avian A, Urschler M, Stollberger R, Olschewski A, Olschewski H, Johnson T, Meinel FG, Bálint Z. Healthy lung vessel morphology derived from thoracic computed tomography. *Frontiers in physiology*. 2018 Apr 10;9:346.
87. Wadensten T, Nystrom E, Franzen K, Stenzelius K, Malmberg L, Samuelsson E. Development and Use of an algorithm For Identifying Women With Urgency Or Mixed Urinary Incontinence Suitable For E-Health Treatment. *Neurourology and Urodynamics*; S5. 2018;37.
88. Alonso-Alonso M, Ziemke F, Magkos F, Barrios FA, Brinkoetter M, Boyd I, Rifkin-Graboi A, Yannakoulia M, Rojas R, Pascual-Leone A, Mantzoros CS. Brain responses to food images during the early and late follicular phase of the menstrual cycle in healthy young women: relation to fasting and feeding. *The American journal of clinical nutrition*. 2011 Aug 1;94(2):377-84.
89. Angeli A, Lainé F, Lavenu A, Ropert M, Lacut K, Gissot V, Sacher-Huvelin S, Jezequel C, Moignet A, Laviolle B, Comets E. Joint model of iron and hepcidin during the menstrual cycle in healthy women. *The AAPS journal*. 2016 Mar;18(2):490-504.
90. Gerassis S, Abad A, Saavedra Á, García JF, Taboada J. Women's Occupational Health: Improving Medical Protocols with Artificial Intelligence Solutions. In *Proceedings of SAI Intelligent Systems Conference 2018* Sep 6 (pp. 1193-1199). Springer, Cham.
91. Brennan-Olsen SL, Solovieva S, Viikari-Juntura E, Ackerman IN, Bowe SJ, et al. Arthritis diagnosis and symptoms are positively associated with specific physical job exposures in lower-and middle-income countries: cross-sectional results from the World Health Organization's Study on global AGEing and adult health (SAGE). *BMC public health*. 2018 Dec;18(1):1-2.
92. Dey A, Hay K, Afroz B, Chandurkar D, Singh K, Dehingia N, Raj A, Silverman JG. Understanding intersections of social determinants of maternal healthcare utilization in Uttar Pradesh, India. *PLoS One*. 2018 Oct 4;13(10):e0204810.
93. Progovac AM, Cook BL, Mullin BO, McDowell A, Sanchez R MJ, Wang Y, Creedon TB, Schuster MA. Identifying gender minority patients' health and health care needs in administrative claims data. *Health Affairs*. 2018 Mar 1;37(3):413-20.
94. Jain A, Kanhangad V. Gender classification in smartphones using gait information. *Expert Systems with Applications*. 2018 Mar 1;93:257-66.
95. Lindenmeyer A, Griffiths F, Green E, Thompson D, Tsouroufli M. Family health narratives: Midlife women's concepts of vulnerability to illness. *Health*. 2008 Jul;12(3):275-93.
96. Lengwiler F, Rappoport D, Jaggi GP, Landau K, Traber GL. Reliability of Cyclotorsion measurements using Scanning Laser Ophthalmoscopy imaging in healthy subjects: the CySLO study. *British Journal of Ophthalmology*. 2018 Apr 1;102(4):535-8.
97. Correia RB, de Araújo Kohler LP, Mattos MM, Rocha LM. City-wide electronic health records reveal gender and age biases in administration of known drug-drug interactions. *NPJ Digital Medicine*. 2019 Jul 23;2(1):1-3.
98. Kweon S, Lee JH, Lee Y, Park YR. Personal health information inference using machine learning on RNA expression data from patients with cancer: algorithm validation study. *Journal of medical Internet research*. 2020;22(8):e18387.
99. Lee JW, Lim HS, Kim DW, Shin SA, Kim J, Yoo B, Cho KH. The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record. *Computer methods and programs in biomedicine*. 2018 Jan 1;153:253-7.

100. Ashokka B, Loh MH, Tan CH, Su LL, Young BE, Lye DC, Biswas A, Illanes SE, Choolani M. Care of the pregnant woman with coronavirus disease 2019 in labor and delivery: anesthesia, emergency cesarean delivery, differential diagnosis in the acutely ill parturient, care of the newborn, and protection of the healthcare personnel. *American journal of obstetrics and gynecology*. 2020 Jul 1;223(1):66-74.
101. Qiu H, Yu HY, Wang LY, Yao Q, et al. Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. *Scientific reports*. 2017 Nov 27;7(1):1-3.
102. Ancochea J, Izquierdo JL, Savana COVID-19 Research Group, Soriano JB. Evidence of gender differences in the diagnosis and management of coronavirus disease 2019 patients: An analysis of electronic health records using natural language processing and machine learning. *Journal of Women's Health*. 2021 Mar 1;30(3):393-404.
103. Akpan IJ, Aguolu OG, Kobara YM, Razavi R, Akpan AA, Shanker M. Association Between What People Learned About COVID-19 Using Web Searches and Their Behavior Toward Public Health Guidelines: Empirical Infodemiology Study. *Journal of Medical Internet Research* 2021;23(9):e28975, 1-19. doi:10.2196/28975; PMID: 34280117.
104. Caballero FF, Soulis G, Engchuan W, Sánchez-Niubó A, Arndt H, Ayuso-Mateos JL, Haro JM, Chatterji S, Panagiotakos DB. Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: the ATHLOS project. *Scientific Reports*. 2017 Mar 10;7(1):1-3.
105. Howell EA, Egorova NN, Janevic T, Balbierz MA, Zeitlin J, Hebert PL. Severe maternal morbidity among Hispanic women in New York City: investigation of health disparities. *Obstetrics and gynecology*. 2017; 129(2):285.
106. Mahunnah M, Taveter K. A scalable multi-agent architecture in environments with limited connectivity: Case study on individualised care for healthy pregnancy. In 2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST) 2013 Jul 24 (pp. 84-89). IEEE.
107. Mailick MR, Movaghar A, Hong J, Greenberg JS, DaWalt LS, Zhou L, Jackson J, Rathouz PJ, Baker MW, Brilliant M, Page D. Health profiles of mosaic versus non-mosaic FMR1 premutation carrier mothers of children with fragile X syndrome. *Frontiers in genetics*. 2018 May 16;9:173.
108. Akpan IJ, Shanker M, Razavi R. Improving the success of simulation projects using 3D visualization and virtual reality. *Journal of The Operational Research Society*, 2020; 71(12), 1900-1926. <https://doi.org/10.1080/01605682.2019.1641649>.
109. Moreira MW, Rodrigues JJ, Marcondes GA, Neto AJ, Kumar N, Diez ID. A preterm birth risk prediction system for mobile health applications based on the support vector machine algorithm. In 2018 IEEE International Conference on Communications (ICC) 2018 May 20 (pp. 1-5). IEEE.
110. Asri H, Mousannif H, Al Moatassime H. Big data analytics in healthcare: case study-miscarriage prediction. *International Journal of Distributed Systems and Technologies (IJDST)*. 2019;10(4):45-58.
111. Baqui A, Ahmed P, Dasgupta SK, Begum N, Rahman M, Islam N, Quaiyum M, Kirkwood B, Edmond K, Shannon C, Newton S. Development and validation of a simplified algorithm for neonatal gestational age assessment—protocol for the Alliance for Maternal Newborn Health Improvement (AMANHI) prospective cohort study. *Journal of global health*. 2017 Dec;7(2).
112. Goin DE, Rudolph KE, Gomez AM, Ahern J. Mediation of Firearm Violence and Preterm Birth by Pregnancy Complications and Health Behaviors: Addressing Structural and Postexposure Confounding. *American journal of epidemiology*. 2020 Aug 1;189(8):820-31.
113. Krishnamurti T, Davis AL, Simhan HN. Worrying yourself sick? Association between pre-eclampsia onset and health-related worry in pregnancy. *Pregnancy hypertension*. 2019 Oct 1;18:55-7.
114. Tayal DK, Meena K, Kumar S. Analysis of various Data Mining Techniques Techniques for Pregnancy related issues and Postnatal health of infant using Machine Learning and Fuzzy Logic. In 2018 3rd International Conference on Communication and Electronics Systems (ICCES) 2018; (789-793). IEEE.
115. de l'Aulnoit AH, Boudet S, Génin M, Gautier PF, Schiro J, de l'Aulnoit DH, Beuscart R. Development of a smart mobile data module for fetal monitoring in E-Healthcare. *Journal of medical systems*. 2018 May;42(5):1-7.
116. Lukkahatai N, Walitt B, Deandrés-Galiana EJ, Fernández-Martínez JL, Saligan LN. A predictive algorithm to identify genes that discriminate individuals with fibromyalgia syndrome diagnosis from healthy controls. *Journal of pain research*. 2018;11:2981.

117. Wiszniewska M, Magnuska J, Lipinska-Ojrzanowska A, Peplonska B, Hanke W, Kalinka J, Skret-Magierlo J, Zadrozny M, Walusiak-Skorupa J. The prophylactic operational model integrated with occupational healthcare-prophylactic of some types of cancers among women/model. *Medycyna pracy*. 2018 Jul 1;69(4):439-56.
118. Chubak J, Onega T, Zhu W, Buist DS, Hubbard RA. An electronic health record-based algorithm to ascertain the date of second breast cancer events. *Medical care*. 2017 Dec 1.
119. Rac MW, Revell PA, Eppes CS. Syphilis during pregnancy: a preventable threat to maternal-fetal health. *American journal of obstetrics and gynecology*. 2017 Apr 1;216(4):352-63.
120. Khamparia A, Gupta D, de Albuquerque VH, Sangaiah AK, Jhaveri RH. Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. *The Journal of Supercomputing*. 2020 Jan 14:1-9.
121. Yu S, Lee JK, Kim JH, Park H, Lee MY, Ryu S, Kwon MJ, Woo HY. Diagnostic performance and establishment of reference limits of HE4 in Korean healthy women. *Gynecologic oncology*. 2016 1;143(1):128-34.
122. Akter S. The gender gap in COVID-19 mortality in the United States. *Feminist Economics*. 2020; 3:1-8.
123. Khan S, Khan A, Maqsood M, Aadil F, Ghazanfar MA. Optimized gabor feature extraction for mass classification using cuckoo search for big data e-healthcare. *Journal of Grid Computing*. 2019 Jun;17(2):239-54.
124. Lynch KE, won Chang J, Matheny ME, Goldfarb A, Efimova O, Coronado G, DuVall SL. Comparison of automated and retrospectively calculated estimated glomerular filtration rate in electronic health record data. *BMC nephrology*. 2018 Dec;19(1):1-0.
125. Haddad SM, Souza RT, Cecatti JG, Barreix M, Tamrat T, Footitt C, Mehl GL, Syah IF, Shankar AH, Tunçalp Ö. Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology. *Journal of Medical Internet Research*. 2020;22(10):e16355.
126. Giardino DL, Huck-Iriart C, Riddick M, Garay A. The endless quarantine: The impact of the COVID-19 outbreak on healthcare workers after three months of mandatory social isolation in Argentina. *Sleep medicine*. 2020 Dec 1;76:16-25.
127. Blotière PO, Weill A, Dalichampt M, Billionnet C, et al. Development of an algorithm to identify pregnancy episodes and related outcomes in health care claims databases: an application to antiepileptic drug use in 4.9 million pregnant women in France. *Pharmacoepidemiology and drug safety*. 2018 Jul;27(7):763-70.
128. Macedonia CR, Johnson CT, Rajapakse I. *Advanced Research and Data Methods in Women's Health*. *Obstetrics & Gynecology*. 2017 Feb 1;129(2):249-64.
129. Shahbaz M, Gao C, Zhai L, Shahzad F, Arshad MR. Moderating effects of gender and resistance to change on the adoption of big data analytics in healthcare. *Complexity*. 2020 Jan 23;2020.