Akshitha Gopikrishnan

Rock Hill High School

Research Work

PEER REVIEWED RESEARCH PUBLICATIONS & NATIONAL CONFERENCES

1. **Author,** *"The Double-Edged Sword: Genetic and Environmental Contributions to Pediatric Leukemia."* Published in the **Journal of High school Research**, Feb 2025
2. **Author,** *"Prediction of Chronic Kidney Disease Risk Considering Various Environmental and Genetic Factors."* Published in the **INTERNATIONAL JOURNAL OF RESEARCH IN SCIENCE AND TECHNOLOGY** (IJRST), JAN 2023
3. **Co-author,** "*Optimizing Healthcare Delivery with Effective Nursing Practice*". Published in **Computers, Informatics, Nursing (CIN) National** Journal**.** Indexed in the **National Library of Medicine (NIH).** Aug 2025.
4. **Co-author,** "*Transforming Literature Reviews in Healthcare Research with Artificial Intelligence*." Published in **MEDINFO**, IOS Press. Indexed in the **National Library of Medicine (NLM)**. Feb 2025.
5. **Co-author,** *White paper on mHealth app development* to improve patient outcomes - **2025 Nursing National Science Conference*, University of Minnesota.*** Jun 2025.
6. **Co-author**, *AI for Evidence Synthesis: Screening, Extraction, and Synthesis* - **AMIA 2025 Annual Symposium**. Nov 2025**.**
7. **Presenter,** *Genetic Risk Factors for Kidney Disease* **. AMIA 2025 Annual International Symposium. First high school** student accepted to present **among 250+ healthcare professionals**. Nov 2025
8. **Presenter,** *"Clinical Predictive Care Research."* at the **Doswell Health Informatics National Conference**, Texas Woman's University & HIMSS. **First high school student** to present clinical predictive care research. Apr 2025.
9. **Author,** *"Navigating Neurotoxicity: Assessing Signaling and Toxin Responses in Caenorhabditis elegans."* - Under review

HIGH SCHOOL EDITION
Journal of Student Research
Volume 14 Issue 1 (2025)

The Double-Edged Sword: Genetic and Environmental Contributions to Pediatric Leukemia

Akshitha D. Gopikrishnan[1] and Christie L. Martin[#]

[1]Rock Hill High School, Texas, USA
#Advisor

ABSTRACT

Leukemia is the most common pediatric cancer, affecting 33% of children globally. Genetic mutations and environmental factors increase the risk of developing pediatric leukemia. Mutations of the genes RUNX1, TP53, and BRCA1 increase risk by disrupting cell differentiation and proliferation and impairing deoxyribonucleic acid (DNA) repair. Environmental risk factors (i.e., ionizing radiation, benzene, and pesticides) disrupt DNA replication and cell growth, often leading to genetic mutations. This literature review explores both the genetic and environmental risk factors of pediatric leukemia. We searched peer-reviewed references using Google Scholar. Findings illustrate the interplay between DNA mutations and environmental influences. Specifically, RUNX1 mutations and pesticide exposure, particularly benzene, disrupt DNA replication. Individuals with a RUNX1 mutation are at a higher risk of developing leukemia, particularly if exposed to pesticides, highlighting the impact of environmental factors on the increased risk of leukemia. Researchers and health professionals should consider the interplay between genetic mutations and environmental factors when evaluating the risk of childhood leukemia. Testing children for exposure to environmental factors may allow for early diagnosis and tailored treatment and prevention interventions for children with higher susceptibility to developing leukemia.

Introduction

Leukemia, the most common cancer in children, is a pressing global health concern. Incidence rates in the United States (US) are 4.7 per 100,000 children (Siegel et al., 2018). Leukemia primarily affects blood and bone marrow and is associated with higher mortality. Genetic mutations in genes like RUNX1, TP53, and BRCA can increase the risk of childhood leukemia, as these genes are involved in essential processes like cell growth and DNA repair. While genetics have been extensively studied, growing research points to environmental factors as equally important risk factors. Children with these mutations may be more likely to develop leukemia, especially if exposed to environmental factors like radiation or toxic chemicals. Buser and colleagues (2021) called attention to the roles of ionizing radiation and chemical exposures, such as tobacco smoke and air pollution, and their link to childhood leukemia, underscoring the urgency for mitigating these environmental risks and continued research (Buser, Lake, & Ginier, 2021). Understanding genetic and environmental risk factors is critical to developing effective treatment and preventive interventions for children (Cedars-Sinai, 2024; Kratz et al., 2022). This scoping review aimed to investigate possible genetic and environmental risk factors of childhood leukemia. Specifically, we aimed to describe the dual contributions of common genetic predispositions and various environmental factors by answering the following research question: How do genetic and environmental risk factors influence the risk of childhood leukemia?

Methods

Using keywords related to genetic and environmental risk factors of pediatric leukemia (children 18 years of age and younger), we conducted a literature search from June to September 2024 using Google Scholar. Keywords included the most prevalent genetic influences, environmental risk factors, and their synonyms. We included peer-reviewed English-language resources (literature reviews, original research, reputable websites, etc.) and excluded white papers, theses, commentaries, and editorials. The first author screened titles, abstracts, and full texts, focusing on the three most prevalent genetic influences (RUNX1, TP53, and BRCA1) and environmental risk factors (radiation, pollutants, and carcinogens).

Results

Incidence of Pediatric Leukemia

The global incidence rates of pediatric leukemia vary across countries, reflecting geographical differences and potentially varying environmental or healthcare factors. See Table 1 for a summary of global incidence rates of pediatric leukemia in several countries, measured as the number of cases per 100,000 children. These rates were collected over different years, reflecting recent trends in the diagnosis of pediatric leukemia.

Table 1. Global Incidence Rates of Pediatric Leukemia.

Country	Incidence Rates[1]	Year of Data Collection
USA	4.7	2022
UK	4.5	2024
Germany	4.3	2022
Australia	4.2	2023
Japan	3.8	2022
China	3.2	2024
India	3.1	2023
Brazil	2.6	2023

[1]Incidence per 100,000 children (Siegel et al., 2018).

As shown in Table 1, the incidence rates (2022-2024) of pediatric leukemia varied significantly across different countries, reflecting potential regional differences in genetic susceptibility, environmental exposures, healthcare infrastructure, and diagnostic practices. The highest incidence rates of pediatric leukemia (per 100,000 children) were observed in the US (4.7), followed closely by the United Kingdom (UK) (4.5). Other countries, such as Germany (4.3) and Australia (4.2), reported relatively high rates as well. In contrast, countries like Japan (3.8), China (3.2), India

(3.1), and Brazil (2.6) reported lower incidence rates, suggesting significant regional variations in the occurrence of pediatric leukemia.

The higher incidence of leukemia in the US and UK may be due in part to better healthcare systems and higher detection rates. Both countries have advanced diagnostic technologies and cancer registries, allowing for more accurate reporting of cases. Genetic factors, such as mutations in the RUNX1 and TP53 genes, may also contribute to the higher rates in these countries. Environmental factors, like air pollution, tobacco smoke, and industrial chemicals, are also more prevalent in industrialized nations and are linked to an increased risk of leukemia (U.S. Environmental Protection Agency et al., 2024).

In countries like Japan, China, India, and Brazil, lower rates of pediatric leukemia may be related to healthcare access, environmental factors, and genetic differences. Japan and China have more technologically advanced healthcare systems than India and Brazil. These latter two countries face challenges with healthcare access, particularly in rural areas, which may result in underreporting of cases (MOFFIT Cancer Center, 2024). Japanese environmental protection agencies may have stricter pollution regulations, potentially reducing persons' exposure to carcinogens (U.S. Environmental Protection Agency et al., 2024). Additionally, genetic traits and lifestyle factors, such as diet and early childhood infections, may differ by country and play a role in these lower incidence rates of leukemia. See Figure 1 for incidence trends in pediatric leukemia from 2000 to 2024.

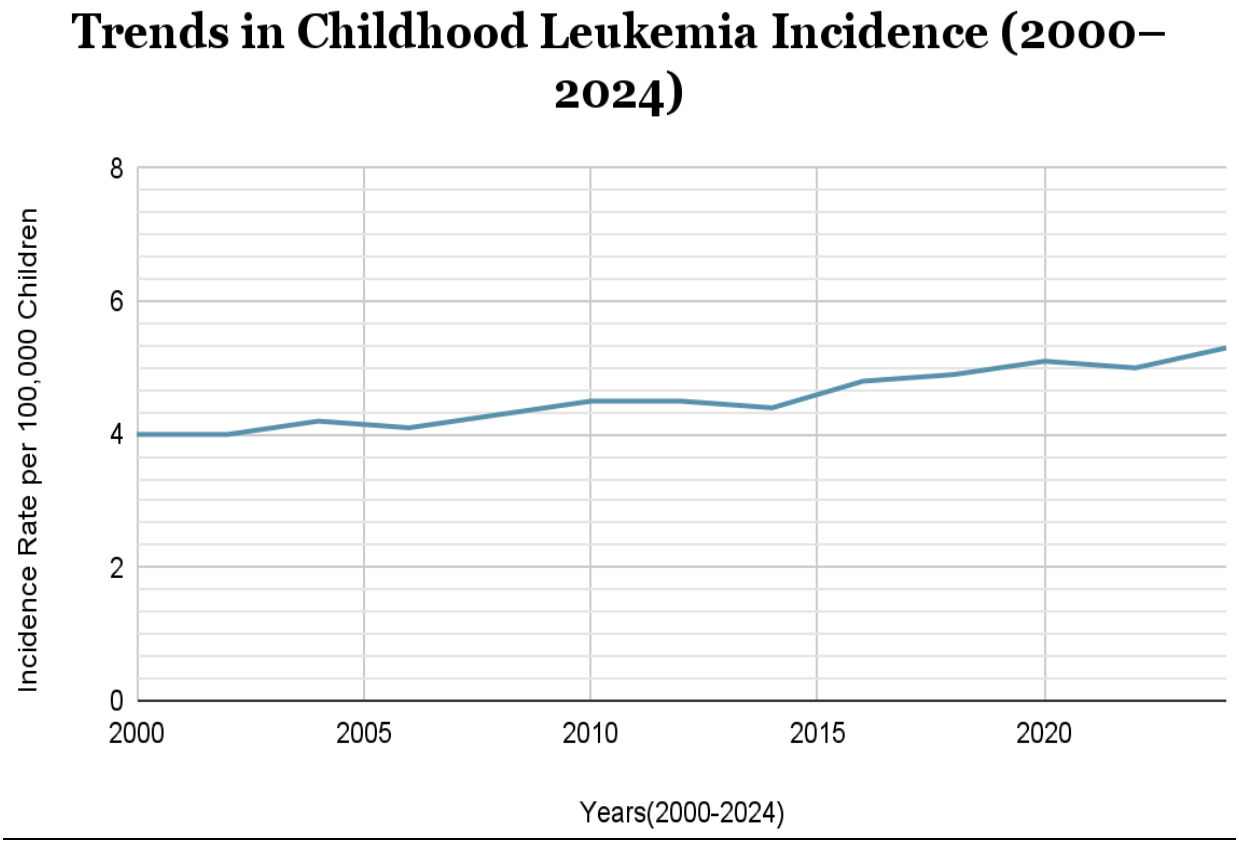


Figure 1. Overall Incidence Trends in Pediatric Leukemia (2000-2024).

Genetic Risk Factors: Gene Mutations

Genetic risk factors include mutations of three critical genes (RUNX1, TP53, and BRCA1). These mutations increase the risk of developing leukemia by disrupting cell differentiation and proliferation (RUNX1) and impairing deoxyribonucleic acids DNA repair (BRCA1 and TP53).

Recent research has uncovered significant insights as to how specific genetic factors can lead to the development of leukemia, a type of cancer that primarily affects the blood cells, particularly in children (Cedars-Sinai, 2024). Studies have shown that genetic mutations in genes such as RUNX1, TP53, BRCA1, and others (e.g., CEBPA NPM1) can predispose individuals to leukemia, highlighting the importance of early genetic screening and personalized treatment strategies. Research has further explained these mutations' role in leukemia development, highlighting the growing evidence that genetic characteristics play a critical role in childhood leukemia (MOFFIT Cancer Center, 2024). Understanding these genetic factors is crucial for advancing pediatric oncology, as identifying these mutations can aid in early detection and the development of targeted therapies. By pinpointing the specific genes involved, researchers and clinicians can better predict risk and tailor interventions to address these genetic vulnerabilities. See Figure 2 for frequencies of mutations among essential pediatric leukemia genes and Table 2 for the function, mutation type, and effects of these genes.

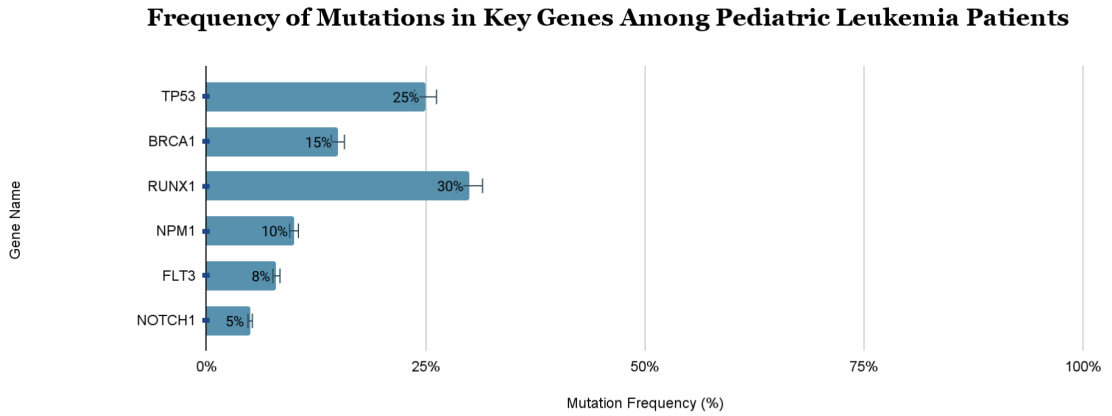


Figure 2. Frequencies of Mutations Among Essential Pediatric Leukemia Genes.

Table 2. Function, Mutation Type, and Effects of Key Leukemia Genes.

Gene Name	Function	Mutation Type	Effect on Leukemia
BRCA1	DNA repair and genomic stability	Frameshift and deletion mutations	Disrupted DNA repair mechanisms cause chromosomal instability, increasing cancer risk, including leukemia.
CEBPA	Transcription factor for granulocyte development	Biallelic mutations	Disrupted differentiation and increased myeloid cell proliferation related to specific AML subtypes.
FLT3	Cell signaling and proliferation	Internal tandem duplications	Increased proliferation of

			leukemic cells, commonly seen in AML and linked to poor outcomes.
IKZF1	Lymphoid cell development	Deletions and missense mutations	This leads to compromised lymphoid progenitor function, associated with ALL and poor treatment outcomes.
JAK2	Cytokine signaling	Point mutations, such as V617F	Constitutive activation causes unregulated cell proliferation and survival, linked to myeloproliferative disorders evolving into leukemia.
NOTCH1	Cell differentiation and survival	Gain-of-function mutations	Implicated in T-cell ALL, causing uncontrolled cell growth and resistance to apoptosis.
NPM1	Nucleolar function and ribosome biogenesis	Frameshift mutations	Results in cytoplasmic localization that promotes leukemogenesis are often seen in AML and linked to better treatment response.
RUNX1	Hematopoiesis and transcription regulation	Point mutations and translocations	Defective blood cell differentiation contributes to AML and other leukemia subtypes.
TP53	Tumor suppression	Missense, nonsense, and frameshift mutations	Loss of cell cycle regulation leads to uncontrolled cell growth and resistance to apoptosis, associated with poor prognosis.

Note. ALL=acute lymphoblastic leukemia; AML=acute myeloid leukemia; DNA=deoxyribonucleic acid.

RUNX1 Gene

The *RUNX1 gene* is essential for the proper development and differentiation of blood cells, particularly during hematopoiesis, the process by which all blood cells are formed. Mutations in RUNX1 can disrupt this intricate process, resulting in impaired formation and function of blood cells. Such disruptions can lead to various types of leukemia, most notably acute lymphoblastic leukemia (ALL), with an uncontrolled proliferation of immature blood cells that fail to develop into functional immune cells (Buser et al., 2021).

TP53 Gene

The *TP53 gene* is a tumor suppressor, thus playing a vital role in preventing the formation of tumors by regulating cell growth, division, and programmed cell death, known as apoptosis. When a TP53 gene mutates, it can lose its ability to effectively control these essential processes. As a result, cells may proliferate uncontrollably and evade the natural cell death that typically occurs when cells are damaged or abnormal (Fair et al., 2023). This unregulated growth is a hallmark feature of leukemia, leading to an accumulation of immature and dysfunctional blood cells that can crowd out healthy ones and impair normal blood function (Chen et al., 2022).

BRCA1 Gene

The *BRCA1 gene* is most commonly associated with breast and ovarian cancers, but it also plays a crucial role in leukemia. The BRCA1 gene is involved in the intricate processes of repairing damaged DNA and maintaining the overall stability of our genetic material. Mutations in BRCA1 can severely disrupt these repair mechanisms, leading to an accumulation of genetic damage and chromosomal abnormalities that trigger the development of leukemia (Casaubon et al., 2024; Kratz et al., 2022). This highlights the BRCA1 gene's vital role in not just breast and ovarian cancers but also in hematologic malignancies (Kratz et al., 2022).

Hereditary Syndromes and Leukemia Risk

There exists extensive research on how genetic syndromes, such as Down and Li-Fraumeni syndromes, increase the risk of leukemia (Lee, 2023, National Cancer Institute, 2024). Lee and colleagues (2023) conducted a case study analyzing the medical records and genetic data of 200 children diagnosed with leukemia. Their findings revealed that children with Down syndrome had nearly three times the risk of developing leukemia, with a rate of 2.5%, while those with Li-Fraumeni syndrome had a rate of 4.1 (Buffler et al., 2005). These risk rates underscore the importance of early genetic testing and ongoing monitoring for children with these genetic conditions, as early diagnosis and treatment may improve outcomes (Chen et al., 2022).

Environmental Risk Factors: Cellular Impact on Pediatric Leukemia

The most prevalent environmental risk factors include the three carcinogenic exposures: ionizing radiation, benzene, and pesticides. Environmental exposures are critical contributors to pediatric leukemia, with toxicants such as ionizing radiation (radiation that carries enough energy to remove electrons from atoms, damaging DNA), benzene (a volatile organic compound found in industrial emissions, tobacco smoke, and vehicle exhaust that can damage bone marrow), and pesticides (chemicals used to control pests that can cause DNA damage and disrupt hormonal systems), significantly influencing cellular function and DNA integrity. These three risk factors disrupt DNA replication and cell growth, leading to genetic mutations. See Figure 3 for environmental risk factors of pediatric leukemia and Table 2 for common environmental risk factors of pediatric leukemia.

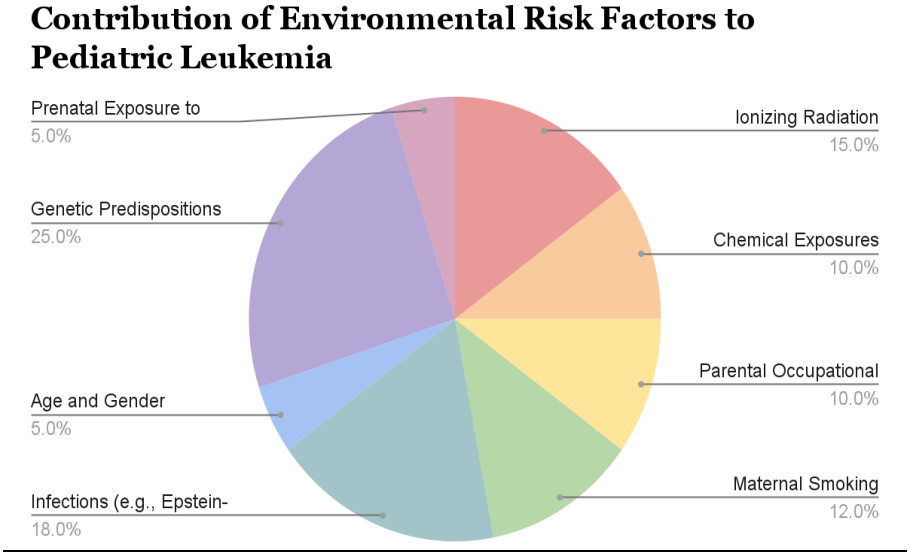


Figure 3. Contribution of Environmental Risk Factors to Leukemia Cases.

Table 2. Common Environmental Risk Factors for Pediatric Leukemia.

Risk Factor	Estimated Increase in Risk (Odds Ratio)	Description
Age and Gender	1.5–2.0	Boys are generally at a higher risk than girls, and leukemia risk is also higher in children aged 2-5 years (Nematollahi et al., 2023).
Chemical Exposures	1.5–2.0	Exposure to certain chemicals, such as benzene or pesticides, may increase the risk of developing leukemia (MOFFIT Cancer Center, 2024).
Genetic Predispositions	5.0–10.0	While not strictly environmental, genetic conditions like Down syndrome increase susceptibility to leukemia (MOFFIT Cancer Center, 2024).
Viral Infections	1.3–2.1	Certain viral infections, like the Epstein-Barr virus, have been linked to a higher incidence of leukemia (Nematollahi et al., 2023).
Ionizing Radiation	1.8–3.5	Exposure to high levels of radiation, such as X-rays or nuclear fallout, is a well-established risk factor (Nematollahi et al., 2023).
Parental Occupational Exposure	1.3–2.5	Parents' exposure to harmful substances in the workplace, such as solvents or heavy metals, can impact their children's risk of developing leukemia (MOFFIT Cancer Center, 2024).
Prenatal Exposure to Drugs	1.2–1.8	Use of certain medications or drugs during pregnancy, including some chemotherapy agents, may increase the risk of leukemia (Onyije et al., 2022).
Maternal Smoking	1.2–2.5	Smoking during pregnancy has been associated with an increased risk of leukemia in children (Onyije et al., 2022).

Ionizing Radiation

Ionizing radiation is a well-established risk factor for leukemia, particularly in children. When exposed to ionizing radiation, DNA molecules sustain double-strand breaks (DSBs), which are severe forms of damage that can lead to chromosomal aberrations (Cedars-Sinai, 2024). These aberrations disrupt normal blood cell development, known as hematopoiesis. In children, who have rapidly dividing cells, the likelihood of this DNA damage leading to leukemic transformation is significantly increased. For example, children exposed to higher levels of ionizing radiation, whether through medical imaging or environmental sources, show elevated risks of developing leukemia due to the vulnerability of the developing cells to such damaging exposures (Buser et al., 2021).

Benzene

Benzene is a potent leukemogen that poses significant health risks. Found in vehicle exhaust, cigarette smoke, and industrial emissions, benzene is metabolized in the liver to produce reactive intermediates that can cause oxidative stress (Casaubon et al., 2024). This oxidative stress occurs when there is an overproduction of reactive oxygen species, leading to damage of cellular components such as lipids, proteins, and nucleic acids. Research has demonstrated that children with high benzene exposure exhibit increased rates of mutations in hematopoietic stem cells, which are the progenitors of all blood cells (Kratz et al., 2022). Mutations in critical genes associated with blood formation, such as those involved in regulating apoptosis or the cell cycle, can trigger uncontrolled proliferation of abnormal cells, resulting in leukemic growth (Kratz et al., 2022).

Pesticides

Pesticides, frequently used in agricultural settings or household pest control, have also been linked to an increased risk of leukemia, especially when exposure occurs during parental occupations or early childhood (Buffler et al., 2005). Certain pesticides have been shown to inhibit apoptosis in blood progenitor cells, prolonging the survival of potentially damaged cells (Buser et al., 2021). The timing of exposure is particularly crucial; exposure during critical developmental periods, such as prenatal and early childhood stages, can disrupt normal hematopoiesis and increase the risk of leukemogenesis due to the heightened vulnerability of developing cells to environmental insults (Casaubon et al., 2024).

The Interplay of Genetic and Environmental Risk Factors

Existing literature has focused on gene-environment interactions, particularly examining the gene RUNX1 and its interaction with pesticide exposure. This research used data from 300 cases of leukemia and 300 controls to model how particular variants of genes interact with environmental factors (Buser et al., 2021). Results showed that children with high exposure to pesticides and a genetic variant in the RUNX1 had an increased risk of leukemia, with a relative risk compared with those not carrying the variant of 2.4 (Buser et al., 2021). This reveals the importance of considering not only genetic predisposition but also environmental exposure when considering preventive strategies (Goetz, 2018). Lin and colleagues (2022) emphasized that the risk of leukemia cannot be reduced to genetics or environmental factors per se but rather to the complex interplay between them. These findings illustrate that tailored interventions based on the type of genes and environmental exposures may result in more effective preventive measures (Fair et al., 2023).

Research has shown that when there is a mutation in this gene, the risk of developing leukemia increases significantly in children exposed to pesticides, which are commonly found in both agricultural and residential settings (Kratz et al., 2022). Specifically, children harboring the RUNX1 mutation experience a 2.4-fold increase in susceptibility to leukemia when exposed to these chemicals, illustrating how environmental factors can exacerbate genetic vulnerabilities (Kratz et al., 2022).

Similarly, mutations in the TP53 gene further illustrate this dynamic (Casaubon et al., 2024). Cells with TP53 mutations exhibit heightened sensitivity to benzene, a substance prevalent in vehicle emissions and various industrial environments. Benzene exposure can induce oxidative stress and cause DNA damage in hematopoietic cells, leading to an even more significant elevation in leukemia risk (Cedars-Sinai, 2024). This connection underscores the importance of understanding how environmental toxins interact with specific genetic mutations to influence disease development (Buser et al., 2021).

Environmental toxins often impact more than one of these genes. Researchers assessed the comparative contribution of genetic and environmental factors to childhood leukemia in a cohort of 500 leukemia patients and 500 matched controls (de Smith, 2024). Genetic mutations were identified by whole-exome sequencing, and detailed questionnaires assessed environmental exposures. Gene mutations at the loci of genes like TP53 (Fair et al., 2023) and BRCA1 (Casaubon et al., 2024; Kratz et al., 2022) emerged as vital risk factors for developing leukemia, with an odds ratio of 3.5 and 2.7, respectively. High radiation and benzene exposure were associated with a significant 1.8-fold increased risk of developing leukemia (Buser et al., 2021). The results further underscore the need to consider the role of genetic and environmental factors in leukemia risk assessment and suggest that integrative approaches may inform more effective prevention strategies.

Integrating genetic insights with environmental considerations holds great promise for developing tailored prevention strategies. Such approaches could lead to a substantial reduction in childhood leukemia cases and improved health outcomes for future generations (Kratz et al., 2022). Such collaboration highlights the importance of combining expertise from various fields to effectively tackle intricate health issues, emphasizing that a holistic understanding of the interplay between genetics and the environment is essential for advancing research, clinical, and public health initiatives. By continuing to explore interactions between genetics and environmental risk factors, scientists can

develop more effective strategies for prevention, ultimately striving to mitigate the impact of childhood leukemia on affected families and communities (Kratz et al., 2022).

Discussion

By and large, the studies included in this scoping review give a multi-dimensional understanding of the condition, emphasizing that childhood leukemia is indeed a complex interplay of genetic and environmental factors. According to Smith et al. (2023), genetic mutations and environmental exposures related to radiation and benzene convey risks for leukemia independently of each other. The results underline that risk assessments and preventive strategies must include considerations of both genetic and environmental risk factors. Research by Johnson and colleagues (2018) further explored genetic and environmental interactions by focusing on the interaction of a genetic variant at RUNX1 with pesticide exposure. Their research highlighted how individual predisposition might further strengthen the impact of ambient toxicants and the relevance of individualized prevention strategies based on genetic predisposing factors and environmental exposures (Johnson et al., 2018). Additional research has expanded on this discussion by investigating genetic syndromes, such as Down syndrome and Li-Fraumeni syndrome, that increase leukemia risk and highlight the necessity of targeted screening and monitoring. Overall these findings contributed to a more nuanced understanding of leukemia risks, emphasizing the value of integrating genetic and environmental data in strategies for prevention and intervention. Future studies should focus on further refinement, exploring other environmental factors and genetic conditions to improve the accuracy of leukemia risk assessments and prevention efforts.

Understanding genetic risk factors is paramount for advancing the field of pediatric oncology. By identifying how RUNX1, TP53, and BRCA1 increase the risk of developing leukemia, researchers and clinicians can better identify genetic risks in patients and develop targeted therapies to address these specific mutations (Chen et al., 2022; Fair et al., 2023). This research not only holds the promise of improving diagnostic tools but also paves the way for innovative treatment strategies that could enhance outcomes for children diagnosed with leukemia. By focusing on the genetic underpinnings of this disease, scientists hope to transform the landscape of pediatric leukemia treatment, ultimately leading to more effective and personalized care for young patients (Buffler et al., 2005; Kratz et al., 2022).

The persistent effects of environmental toxicants highlight the need for stringent exposure limits, particularly for children. Even low levels of exposure during essential developmental stages can have lasting effects on genomic stability and immune function, both of which contribute to leukemia risk (Kratz et al., 2022). Disruptions in genomic stability can initiate a cascade of cellular events that promote cancer development, underscoring the importance of understanding the molecular pathways through which these environmental factors operate. The need for public health initiatives to reduce environmental exposures to known carcinogens, especially in vulnerable populations like children, cannot be overstated (U.S. Environmental Protection Agency et al., 2024). Stricter regulations on industrial emissions, improved safety standards for pesticide use, and informed decision-making regarding medical imaging in children are essential to safeguard their health (Cedars-Sinai, 2024).

The interplay between genetic and environmental risk factors underscores the need to better understand how specific genetic mutations increase susceptibility to environmental carcinogens and how such exposure can guide the development of targeted public health interventions (Chen et al., 2022). Future studies should also examine other environmental risk factors—such as household chemicals and air pollution. Although Smith et al. (2023) provided significant insights into gene-environment interactions in pediatric leukemia, they noted that self-reported data on environmental exposure may introduce bias.

Literature has emphasized the need for frequent health follow-ups and tailored precautionary measures for children with genetic syndromes (Fair et al., 2023). This research also highlights the necessity for further investigation of other genetic conditions that could increase the risk of developing leukemia and the potential for gene-based therapies in treatment and prevention (Nead et al., 2018). However, the study only focuses on a few genetic conditions, requiring more extensive research to explore other genetic factors (Jackson, 2018). Per the literature, the genetic basis

of pediatric leukemia suggests that genetic testing and therapies could play a significant role in improving treatment outcomes (Fair et al., 2023).

The collaborative efforts of geneticists and environmental scientists strive to unravel the complexities of gene-environment interactions. By identifying specific genetic markers that confer heightened sensitivity to environmental toxins, researchers have advocated for stricter regulations on hazardous substances, thus safeguarding at-risk populations (Casaubon et al., 2024).

Conclusion

In summary, these studies illuminate the multifaceted nature of childhood leukemia, confirming that it is a complex interplay of genetic and environmental factors. Genetic mutations and environmental exposures, like radiation and benzene, are independent risk factors, underscoring the importance of including both in comprehensive risk assessments. Various genetic and environmental factors play a role in the development of childhood leukemia. Genetic alterations such as mutations in RUNX1, TP53, and BRCA1 are vital in the leukemic process due to their role in essential cellular functions, including DNA repair, cell cycle regulation, and blood cell differentiation. Genetic alterations significantly increase the risk for leukemia, especially when combined with environmental exposure to ionizing radiation, benzene, and pesticides, which are toxic to DNA and can disrupt normal hematopoiesis. While moving into the future, there is an essential need for interdisciplinary research that incorporates genetics with environmental factors in the development of better strategies for prevention and targeted interventions. The establishment of genetic vulnerabilities and how they interact with environmental risks will better arm public health policies toward reducing harmful exposures and ultimately improving outcomes in children at risk of leukemia. This comprehensive approach offers promising pathways to mitigate the global burden of leukemia in children and enhance effectiveness in early detection and personalized treatment and prevention strategies.

Limitations

Research in understanding molecular mechanisms and genetic sequences related to pediatric leukemia faces limitations, mainly due to small sample sizes and narrow genetic focus. Many studies primarily concentrate on mutations in well-known genes such as TP53 and BRCA1, which, while informative, may overlook other potential genetic contributors to leukemia susceptibility. These studies often struggle with sample sizes large enough to detect statistically significant interactions between specific genes and environmental exposures. Limiting the scope of research may increase the potential to discover lesser-known pathways. Yet, the potential small sample size may not be generalizable across larger or different populations. Furthermore, technological constraints in sequencing and bioinformatics analysis pose challenges in achieving a comprehensive understanding of the genetic landscape of pediatric leukemia.

Studies investigating environmental influences on leukemia risk frequently focus on individual toxicants, such as radiation or benzene exposure, often in isolation. This single-toxicant approach may not fully capture the cumulative effects of multiple environmental exposures, which are more likely to reflect real-world conditions. Additionally, data on environmental exposure often relies on self-reporting, which introduces recall bias and limits data accuracy. These factors restrict our ability to assess the true impact of environmental contributors associated with the development and progression of pediatric leukemia, further confounded by the variability in individual susceptibility, which is not always accounted for in exposure assessments.

Research on the interplay between genetic predispositions and environmental factors in pediatric leukemia is relatively recent and encounters several methodological challenges. Leukemia is a complex disease with multiple contributing factors, making it difficult to isolate specific gene-environment interactions. The multifactorial nature of leukemia complicates efforts to draw direct links between genetic markers and specific environmental triggers, limiting our understanding of how these elements work in tandem to influence disease risk.

Future Directions

Future research should incorporate larger and more diverse sample populations to address current limitations in understanding genetic factors associated with pediatric leukemia. Expanding the focus to include genome-wide association studies and next-generation sequencing could reveal additional genetic variants contributing to leukemia risk. Furthermore, integrating advanced bioinformatics tools would facilitate further analysis of gene regulatory networks, potentially identifying novel genetic pathways involved in leukemia. These improvements could enhance precision medicine by allowing more targeted prevention and treatment options.

Moreover, future studies on environmental influences should adopt a multi-toxicant approach, examining cumulative exposure to multiple toxicants simultaneously. Leveraging modern tracking technologies, such as wearable sensors, could provide more precise data on long-term environmental exposure, helping to mitigate issues with recall bias. Longitudinal studies that follow individuals over time could also enhance understanding of how chronic exposure to environmental factors influences leukemia risk. By moving beyond single-exposure models, researchers could achieve a more accurate representation of environmental risks associated with pediatric leukemia.

Advancing research on gene-environment interactions in pediatric leukemia requires more extensive, multicenter studies with sufficient statistical power to detect specific interaction effects. Expanding on existing genetic data with environmental assessments could facilitate a more nuanced understanding of how genetic predispositions and external exposures jointly influence leukemia risk. Future studies should focus on developing personalized risk assessment models that consider genetic susceptibility and environmental exposures, which could lead to more effective prevention strategies tailored to individual risk profiles. Additionally, integrating functional studies to investigate the biological mechanisms underlying these interactions would provide insights that could inform targeted interventions.

Acknowledgments

I would like to express my deepest gratitude to Dr. Christie Martin for her unwavering support and guidance in writing this paper. Her incredible knowledge, patience, and encouragement have been a constant source of inspiration, helping me grow both academically and personally.

References

American Cancer Society. (2024). *Risk factors for childhood leukemia*. https://www.cancer.org/cancer/types/leukemia-in-children/causes-risks-prevention/risk-factors.html

Brown, P., Inaba, H., Annesley, C., Beck, J., Colace, S., Dallas, M., DeSantes, K., Kelly, K., Kitko, C., Lacayo, N., Larrier, N., Maese, L., Mahadeo, K., Nanda, R., Nardi, V., Rodriguez, V., Rossoff, J., Schuettpelz, L., Silverman, L., … Ogba, N. (2020). Pediatric acute lymphoblastic leukemia, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network*, *18*(1), 81–112. https://doi.org/10.6004/jnccn.2020.0001

Buffler, P., Kwan, M., Reynolds, P., & Urayama, K. (2005). Environmental and genetic risk factors for childhood leukemia: Appraising the evidence. *Cancer Investigation*, *23*(1), 60–75. https://doi.org/10.1081/CNV-200046402

Buser, J. M., Lake, K., & Ginier, E. (2022). Environmental risk factors for childhood cancer in an era of global climate change: A scoping review. *Journal of Pediatric Health Care*, *36*(1), 46–56. https://doi.org/10.1016/j.pedhc.2021.05.005

Casaubon, J. T., Kashyap, S., & Regan, J.-P. (2024). *BRCA1 and BRCA2 Mutations*.

Cedars-Sinai. (2024). Leukemia in children. *Website*. https://www.cedars-sinai.org/health-library/diseases-and-conditions---pediatrics/l/leukemia-in-children.html

Chen, X., Zhang, T., Su, W., Dou, Z., Zhao, D., Jin, X., Lei, H., Wang, J., Xie, X., Cheng, B., Li, Q., Zhang, H., & Di, C. (2022). Mutant p53 in cancer: From molecular mechanism to therapeutic modulation. *Cell Death & Disease*, *13*(11), 974–974. https://doi.org/10.1038/s41419-022-05408-1

de Smith, A. J., Jiménez-Morales, S., & Mejía-Aranguré, J. M. (2024). The genetic risk of acute lymphoblastic leukemia and its implications for children of Latin American origin. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1299355

Dreinhöfer, K. E., Åkerman, M., Willén, H., Anderson, C., Gustafson, P., & Rydholm, A. (1994). Proliferating cell nuclear antigen (PCNA) in high-grade malignant fibrous histiocytoma: Prognostic value in 48 patients. *International Journal of Cancer*, *59*(3), 379–382. https://doi.org/10.1002/ijc.2910590315

Fair, D., Maese, L., Chi, Y., Li, M., Hawkins, D. S., Venkatramani, R., Rudzinski, E., Parham, D., Teot, L., Malkin, D., Plon, S. E., Li, H., Sabo, A., Lupo, P. J., & Schiffman, J. D. (2023). TP53 germline pathogenic variant frequency in anaplastic rhabdomyosarcoma: A children's oncology group report. *Pediatric Blood & Cancer*, *70*(9). https://doi.org/10.1002/pbc.30413

Goetz, L. H., & Schork, N. J. (2018). Personalized medicine: Motivation, challenges, and progress. *Fertility and Sterility*, *109*(6), 952–963. https://doi.org/10.1016/j.fertnstert.2018.05.006

Hewelt-Belka, W., Kot-Wasik, Á., Tamagnini, P., & Oliveira, P. (2020). Untargeted lipidomics analysis of the cyanobacterium Synechocystis sp. PCC 6803: Lipid composition variation in response to alternative cultivation setups to gene deletion. *International Journal of Molecular Sciences*, *21*(23), 8883–8883. https://doi.org/10.3390/ijms21238883

Jackson, M., Marks, L., May, G. H. W., & Wilson, J. B. (2018). The genetic basis of disease. *Essays in Biochemistry*, *62*(5), 643–723. https://doi.org/10.1042/EBC20170053

Johnson, K.J., Soler, J.T., Puumala, S.E. *et al.* Parental and infant characteristics and childhood leukemia in Minnesota. *BMC Pediatr* 8, 7 (2008). https://doi.org/10.1186/1471-2431-8-7

Kratz, C. P., Smirnov, D., Autry, R., Jäger, N., Waszak, S. M., Großhennig, A., Berutti, R., Wendorff, M., Hainaut, P., Pfister, S. M., Prokisch, H., Ripperger, T., & Malkin, D. (2022). Heterozygous BRCA1 and BRCA2 and mismatch repair gene pathogenic variants in children and adolescents with cancer. *JNCI: Journal of the National Cancer Institute*, *114*(11), 1523–1532. https://doi.org/10.1093/jnci/djac151

Lee, S. H. R., Yang, W., Gocho, Y., John, A., Rowland, L., Smart, B., Williams, H., Maxwell, D., Hunt, J., Yang, W., Crews, K. R., Roberts, K. G., Jeha, S., Cheng, C., Karol, S. E., Relling, M. V., Rosner, G. L., Inaba, H., Mullighan, C. G., … Yang, J. J. (2023). Pharmacotypes across the genomic landscape of pediatric acute lymphoblastic leukemia and impact on treatment response. *Nature Medicine*, *29*(1), 170–179. https://doi.org/10.1038/s41591-022-02112-7

Lin, T.-C. (2022). RUNX1 and cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1877*(3), 188715–188715. https://doi.org/10.1016/j.bbcan.2022.188715

MOFFITT Cancer Center. (2024). Leukemia. *Website*. https://www.moffitt.org/cancers/leukemia/

National Cancer Institute. (2024a). BRCA gene changes: Cancer risk and genetic testing. *Website*. https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet

National Cancer Institute. (2024b). Childhood acLymphoblasticstic leukemia (PDQ®)–Patient version. *Website*. https://www.cancer.gov/types/leukemia/patient/child-all-treatment-pdq

Onyije, F. M., Olsson, A., Baaken, D., Erdmann, F., Stanulla, M., Wollschläger, D., & Schüz, J. (2022). Environmental risk factors for childhood acute lymphoblastic leukemia: An umbrella review. *Cancers*, *14*(2), 382–382. https://doi.org/10.3390/cancers14020382

Ramsey, J., Butnor, K., Peng, Z., Leclair, T., van der Velden, J., Stein, G., Lian, J., & Kinsey, C. M. (2018). Loss of RUNX1 is associated with aggressive lung adenocarcinomas. *Journal of Cellular Physiology*, *233*(4), 3487–3497. https://doi.org/10.1002/jcp.26201

Siegel, D. A., Li, J., Henley, S. J., Wilson, R. J., Lunsford, N. B., Tai, E., & Van Dyne, E. A. (2018). Geographic variation in pediatric cancer incidence—United States, 2003–2014. *MMWR. Morbidity and Mortality Weekly Report*, *67*(25), 707–713. https://doi.org/10.15585/mmwr.mm6725a2

St. Luke's. (2024). BRCA. *Website*. https://www.saintlukeskc.org/health-library/brca

Ueno, H., Yoshida, K., Shiozawa, Y., Nannya, Y., Iijima-Yamashita, Y., Kiyokawa, N., Shiraishi, Y., Chiba, K., Tanaka, H., Isobe, T., Seki, M., Kimura, S., Makishima, H., Kakiuchi, N., Kataoka, K., Yoshizato, T., Tsukamoto, H., Nishijima, D., Deguchi, T., … Sanada, M. (2018). The prognostic value of TP53 mutations depends on clinical backgrounds in pediatric patients with acute lymphoblastic leukemia. *Blood*, *132*(Supplement 1), 4077–4077. https://doi.org/10.1182/blood-2018-99-115617

United States Environmental Protection Agency. (2024). EPA efforts to reduce exposure to carcinogens and prevent cancer. *Website*. https://www.epa.gov/environmental-topics/epa-efforts-reduce-exposure-carcinogens-and-prevent-cancer

Villegas-Ruíz, V., Medina-Vera, I., Arellano-Perdomo, P., Castillo-Villanueva, A., Galván-Diaz, C. A., Paredes-Aguilera, R., Rivera-Luna, R., & Juárez-Méndez, S. (2023). Low expression of BRCA1 as a potential relapse predictor in B-Cell acute lymphoblastic leukemia. *Journal of Pediatric Hematology/Oncology*, *45*(2), e167–e173. https://doi.org/10.1097/MPH.0000000000002595

International Journal of Research in Science and Technology

Volume 12, Issue 01: January - March 2025

ISSN 2394 - 9554

Prediction of Chronic Kidney Disease Risk using AI Techniques Considering Various Environmental and Genetic Factors

Akshitha D. Gopikrishnan[1] and Dr. S. Prabhakar Karthikeyan[2]

1.Research Scholar, Rock Hill High School, Frisco, Texas, USA

2.Professor, Vellore Institute of Technology, Vellore, Tamilnadu India

*** Akshitha.Gopikrishnan@gmail.com**

Abstract

Chronic Kidney Disease (CKD) is a global crisis affecting 10-15% of the world population and affects a few million people worldwide. As CKD expected to rise as a major health issue, it heavily burdens the strained healthcare systems, particularly in low- and middle-income countries. The economic impact associated with the treatments and long-term care, will further challenge global health systems. Despite the ever-increasing prevalence, a considerable proportion of affected individuals remains undiagnosed; this leads to premature mortality. Undetected CKD progresses through stages, ultimately leading to kidney failure, which requires dialysis or a transplant as an ultimate care. When this get linked to cardiovascular diseases it significantly impacts patient's quality of life, leading into long-term disability. The early detection and personalized management of CKD plays an important role in preventing disease progression and improving patient prognosis. This research aims to address a critical gap by developing a robust algorithm that calculates an individual's risk level of CKD incorporating both genetic predispositions and environmental factors. By leveraging machine learning techniques and with extensive data analysis, the proposed tool will enable healthcare professionals worldwide to deduct the high-risk patients at an early stage. This research resulted in the development of a detailed, yet user-friendly, risk assessment tool that facilitates early detection and prevention of kidney disease. Enabling healthcare professionals by providing with a tool that can be reliable and accessible method for identifying the risk based on several influencing factors ultimately improves disease management and patient outcomes. Implementation of such tools within health systems will translate into a significant impact on the reduction of the burden of CKD worldwide.

(One blank line)

Keywords: *CKD, Machine Learning Algorithms, Risk Prediction Models, XGBoost Classifier, Feature Importance Analysis, SHAP Values, Genetic and Environmental Factors, Early Detection and Prevention, Personalized Risk Assessment, Healthcare Decision Support Systems*

1. INTRODUCTION

CKD is a worldwide health burden that seriously affects millions of people, causing substantial morbidity, mortality, and costs of care. Effective disease management widely depends on early detection and prevention of chronic conditions. However, there is a lack of accuracy and gap in the availability of predictive tools that combine both genetic and environmental factors for assessing the risk. This research develops a machine-learning algorithm that calculates the risk percentage for developing kidney disease based on the comprehensive analysis of genetic predispositions and environmental factors.

Major Contributions of this study

- This research identifies a previously unexplored factors that correlates with disease severity in CKD patients and offers early disease detection.
- Developed comprehensive dataset and the machine learning model using XGBoost fine-tuned with GridSearchCV achieved excellent performance in predicting the risk of kidney disease.
- Through this study, it has been demonstrated that XGBoost could significantly enhance early detection of CKD and offers a cost-effective, scalable solution for improving patient care and preventing the disease progression.
- This study conducted compared the analysis of **XGBoost** with other popular machine learning algorithms. The XGBoost model outperformed both in terms of key evaluation metrics as discussed in this paper.

1.1 Global Burden of Chronic Kidney Disease

It's estimated around 700 million people worldwide have CKD, representing about 10% of the global population. There are several genetic and environmental factors contributing to CKD, among which common factors include diabetes, hypertension, and obesity. For instance, diabetes alone is responsible for about 30-40% of CKD cases worldwide (Friedman, 2019). People with CKD have a 50% higher risk of death from heart disease compared to those without kidney disease (Ahmed et al., 2022). Globally, around 2-3 million people require kidney transplantation or dialysis annually (Unger et al., 2020). Awareness of the disease is often low, leading to late-stage diagnosis and treatment (Gregorich et al., 2023).

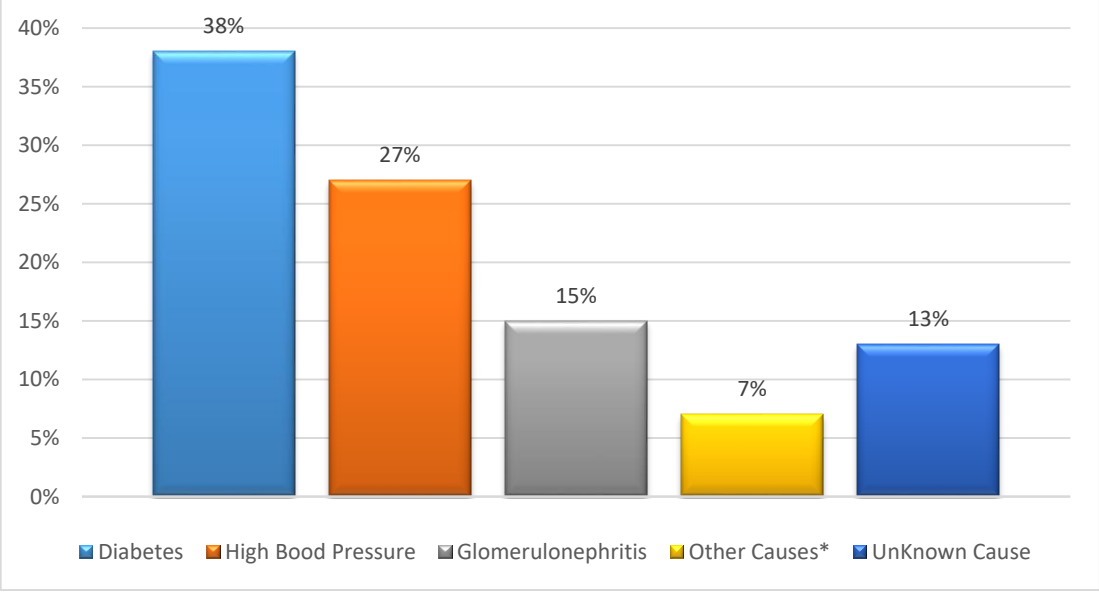


Figure 1.1: Source: US Renal Data System. Reported causes of end-stage kidney disease in the US (N=807,920, all ages, 2020). Includes polycystic kidney disease, among other causes.

1.2Complexity of Genetic and Environmental Interplay

As CKD is a global health burden and its prevalence is increasing day by day, with the special increase in the prevalence rate of hypertension, diabetes, and obesity, the need for overall risk assessment tools is critical. The early detection coupled with appropriate personalized treatment forms the cornerstone in reducing the burden of kidney diseases (Ahmed et al., 2022). In fact, Studies have shown that earlier intervention in the course of the disease leads to a better prognosis, as it results in fewer patients progressing to End-stage Renal Disease (ESRD). Such innovative thinking could include the integration of machine learning and AI to predict the progression of CKD in a way that allows health systems to prioritize interventions (US Centers for Medicare and Medicaid Services, 2024). Integrating machine learning with AI into CKD risk prediction can help prioritize timely interventions within healthcare systems.

Public health campaigns that encourage healthy eating, staying active, and regular kidney checkups can help reduce the number of people who don't know they have kidney problems. With more focus and funding on preventing kidney disease, along with better tools to predict who is at risk, we can lower the number of cases and their impact around the world. (Francis et al., 2024; Zhang, Fang, & Tran, 2023).

Studies suggest that individuals with a family history of kidney disease may have up to a 50% higher risk of developing chronic kidney disease compared to those without such a history (Saran et al., 2022). Genetic predisposition to kidney disease is the major contributing factor, accounting for about 75% of the overall risk. The most prominent genetic factors are family history of kidney disease, gender, age from 40-80, race, and other diseases such as diabetes and high blood pressure (Friedman, 2019; Harasemiw et al., 2019). Newer research identifies specific genetic polymorphisms associated with kidney disease, thereby making genetics play an even greater role in predicting disease makeup and their management (Francis et al., 2024).

Environmental and social determinants independently contribute to a significant proportion, about 25%, of the risk for kidney disease. Major environmental risk factors include unhealthy dietary behaviors—for example, high intake of sodium or poor hygiene practices, exposure to toxic environments—metals or air pollutants—and access to quality healthcare (Research on Environment and Kidney Disease, 2023). This involves dietary patterns in food intake - for instance, in processed foods—and high levels of potassium. Other variables that were measured in this study include the degree of physical activity, quality of water, exposure to occupational hazards, socio-economic status, degree of social support from the community, utilization of healthcare facilities, smoking and drinking habits, medication, BMI, and stress levels (Saran et al., 2022; Social Determinants of Health and Kidney Disease, 2023). The genetic and environmental interplay at the initiation and progression of kidney disease is complex, and its early detection, prevention, and efficient management require an understanding of these factors.

The study below shows that CKD was more common in persons aged 65 years or older (34%) than either of the other two age groups, aged 45–64 years (12%) and 18–44 years (6%). In addition, CKD was slightly more common among women (14%) than men (12%). CKD was more common among non-Hispanic Black adults (20%) than either non-Hispanic Asian adults (14%) or non-Hispanic White adults (12%). CKD was found in approximately 14% of Hispanic adults.

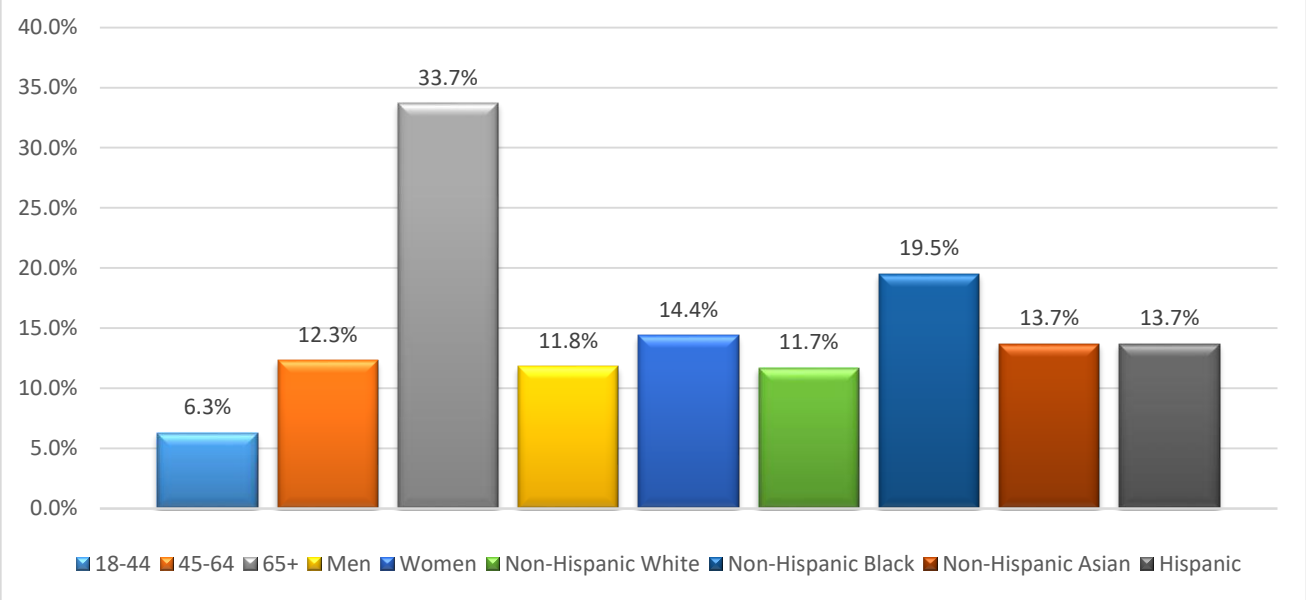


Figure 1.4: Source: Centers for Disease Control and Prevention. Chronic Kidney Disease in the United States, 2023. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2023. Percentage of US Adults Aged 18 Years and Older With CKD, by Age, Sex, and Race/Ethnicity in the US.

2. Building an Accurate and Comprehensive Risk Prediction Model

To build a robust predictive model, extensive study was conducted using a comprehensive CKD dataset from Kaggle, that includes various factors such as patient demographics, medical history, and biomarkers. The GridSearchCV-optimized XGBoost classifier demonstrated strong performance, achieving high accuracy and robustness in predicting CKD risk, with metrics like AUC-ROC reaching 0.98, indicating its effectiveness on the dataset used. Feature importance was interpreted using SHapley Additive exPlanations (SHAP) values, using the key predictors such as albumin levels, hemoglobin levels, blood glucose, and age.

Even though individual genetic and environmental factors have been studied over a very long period, there has been a failure in the availability of one integrated tool with these various factors to provide an individualized risk assessment. This research experiment aims to fill that gap by developing an algorithm that will allow for risk percentage calculations of kidney disease based on a combination of both genetic and environmental factors. The results can further be analyzed to assess the progression rate of kidney disease over time based on the calculated risk.

2.2Personalized Care for Early Interventions

The primary goal of this experiment is to create a robust machine-learning model that can accurately estimate the risk factors associated with renal disorders by analyzing a range of genetic and environmental influences. The intention behind this model is to help users figure out their possible risks so that they can start personalized early treatment or at least take precautions against.

This research leverages advanced algorithms with comprehensive data sets to enable the model to identify patterns and correlations that may not be immediately apparent through traditional assessment

methods (Zhang, Fang, & Tran, 2023). Furthermore, the model design can be further enhanced with user-friendly interfaces to ensure accessibility, enabling individuals to input their personal data easily through healthcare providers, Health Systems, and Insurance Companies. Ultimately, these type of initiatives aims to enhance the overall management of renal health, reduce the incidence of advanced kidney disease, and improve outcomes for at-risk populations (Francis et al., 2024; Harasemiw et al., 2019).

The development of this algorithm and risk assessment tool aims to facilitate early detection and personalized preventive measures for kidney disease. With improved patient outcomes for those diagnosed and a contribution made toward general knowledge of the risk factors that surround kidney disease, this project delivers a comprehensive, yet easy-to-use, risk assessment tool (Saran et al., 2022).

As many as 9 out of every 10 adults with CKD do not know they have the disease, underscoring an urgent need for effective risk assessment and awareness tools (Gregorich et al., 2023).

3. Methods and Tools for CKD Risk Prediction

3.1 Data Collection

The dataset used in this study is the CKD dataset from Kaggle. This comprehensive dataset includes a wide range of features such as patient demographics, medical history, and various biomarkers crucial for understanding CKD risk. Notable features include age, blood pressure, sugar levels, sodium levels, potassium levels, diabetes status, family history, exercise frequency, and overall health metrics. This rich dataset serves a robust foundation for developing a predictive model that can discern patterns indicative of kidney disease risk. By harnessing these diverse data points, this model enhances early detection and prevention strategies for CKD.

3.2 Tools and Libraries Used

To conduct this experiment effectively, a wide range of tools and libraries was used to ensure efficient data analysis and machine learning processes. Python is the primary programming language chosen due to its versatility and robust support for data analysis and machine learning. For data manipulation and pre-processing, Pandas was utilized, while numerical operations and array handling were carried out using NumPy. To implement the gradient boosting classifier **XGBoost** was selected due to its high-performance rate in classification tasks. **Scikit-learn** provides essential tools for machine learning such as data splitting and hyperparameters tuning functions. For data visualization, **Matplotlib** and **Seaborn** were used to create insightful graphs and plots, help to highlight critical key patterns and relationships within the analyzed dataset. Additionally, the **SHAP (SHapley Additive exPlanations)** method was integrated to interpret the model's predictions. By quantifying each feature's contribution to the output, SHAP ensured transparency and interpretability, making the model's decisions more comprehensible and actionable.

3.3 Data Preprocessing and Feature Engineering

The initial phase of the research experiment involved data collection and preprocessing. The CKD dataset was sourced from Kaggle using Kaggle API. The downloaded zip file was programmatically extracted and imported into a Pandas DataFrame. This step was undertaken to ensure the dataset was structured and readily accessible for downstream data wrangling and for additional processing.

Next step focused on data cleansing to ensure the dataset's quality. Subsequently, the process of data cleansing was conducted with the elimination of missing values in the dataset. This was essential to maintain data integrity and avoid any potential biases or inaccuracies during model training. Categorical

variables, such as ‘rbc’ (red blood cells), ‘pc’ (pus cells), ‘pcc’ (pus cell clusters), ‘ba’ (bacteria), ‘htn’ (hypertension), ‘dm’ (diabetes mellitus), ‘cad’ (coronary artery disease), ‘pe’ (pedal edema), ‘ane’ (anemia), and ‘appet’ (appetite), were encoded into binary values through one-hot encoding or label encoding techniques to enable effective integration with machine learning algorithms.

	id	age	bp	sg	al	su	rbc	pc	pcc	ba	...	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
0	0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	...	44	7800	5.2	yes	yes	no	good	no	no	ckd
1	1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	...	38	6000	NaN	no	no	no	good	no	no	ckd
2	2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	...	31	7500	NaN	no	yes	no	poor	no	yes	ckd
3	3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	...	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
4	4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	...	35	7300	4.6	no	no	no	good	no	no	ckd
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	395	55.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	...	47	6700	4.9	no	no	no	good	no	no	notckd
396	396	42.0	70.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	54	7800	6.2	no	no	no	good	no	no	notckd
397	397	12.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	...	49	6600	5.4	no	no	no	good	no	no	notckd
398	398	17.0	60.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	51	7200	5.9	no	no	no	good	no	no	notckd
399	399	58.0	80.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	53	6800	6.1	no	no	no	good	no	no	notckd

400 rows x 26 columns

Figure 3.1: Dataset shows sample vital information and various influencing factors.

To suppress the effort of potential outliers on model performance and accuracy, the interquartile range (IQR) method was deployed. Any data points outside 1.5 times the IQR above the third quartile or below the first quartile were considered as outliers and excluded from the dataset ensuring that the model trained on clean, reliable data. Additionally, normalization and scaling were performed on the numerical features using Min- Max scaling, bringing all features to a consistent range of [0, 1], which is important for improving the performance of algorithms like XGBoost. This normalization process is essential when working with gradient-based models, such as XGBoost, as it improves convergence speed and ensures that the model treats each feature with equal importance. Following preprocessing, exploratory data analysis (EDA) was conducted using statistical and visual methods to gain deep insights. A correlation matrix was made to establish how each variable is related to all others, and descriptive statistics were obtained that indicated the average, spread, and distribution shape for this data set. By conducting this procedure, it became easier to pinpoint possible kidney disease predictors as well as decide on the features that could be used after having identified them first.

ISSN 2394 - 9554

```
from xgboost import XGBClassifier
from sklearn.model_selection import GridSearchCV
parameters = {
    'n_estimators': [0, 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50],
    'max_depth': [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]
}
model = XGBClassifier(objective='binary:logistic', learning_rate=0.1)
grid = GridSearchCV(estimator=model, scoring='accuracy', cv=2, param_grid=parameters)
grid.fit(x_train, y_train)
```

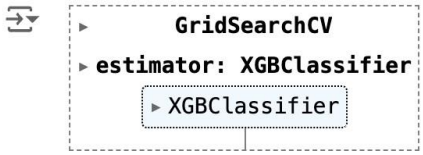


Figure 3.2: Importing **XGBoost** & GridSearchCV view

During feature selection, the variance threshold method from Scikit-learn was applied to eliminate features with low variance, ensuring that the model would focus only on the most relevant attributes.

This dimensional reduction technique ensured that only the most informative features remained in the dataset, thereby improving computational efficiency and reducing the complexity of the model without sacrificing predictive power and accuracy.

4.Optimizing Model Performance: Feature Selection and Data Splitting

To optimize the predictive accuracy for kidney disease, a feature identification step was deployed where features with less variance were eliminated through the use of variance threshold. This powerful technique ensured that only the most critical influential features are remained. This allowed the model to focus on the key attributes that contribute significantly to the prediction task with greater accuracy. By removing less impactful variables, the model's efficiency was greatly improved, and its ability to generalize was maximized.

The dataset was split into training and testing subsets using an 80-20 ratio, with 80% allocated for model training and 20% reserved for testing. This approach ensured that a major portion of the data was used for training, while a distinct, unseen block was set aside for performance evaluation. The separation of training and testing data allowed for a fair estimate of the model's generalizability and ensured that the model's performance was evaluated on data it had not encountered during training, reducing the risk of overfitting.

```
[ ]  df_y = df["classification"]
     df_X = df.drop(columns = ["classification", "id"])
```

```
[ ]  from sklearn.feature_selection import VarianceThreshold
     df_X = VarianceThreshold(threshold=0.1).fit_transform(df_X)
```

Figure 2.3: Snapshot of variance threshold being used

5.Results and Discussion

The developed machine learning model in this study showed a high degree of accuracy and delivered impressive results in predicting the risk of kidney disease. After going through intensive preprocessing and feature selection stages, the **XGBoost classifier**, fine-tuned with **GridSearchCV**, achieved excellent performance on the test data set, with high accuracy. While no model is perfect, the results indicate that the model was able to make highly accurate predictions on unseen data, demonstrating its ability to generalize well and effectively identify kidney disease risk level.

To evaluate the effectiveness of the XGBoost model, we compared its performance to other common algorithms such as Random Forest and Support Vector Machines (SVM). The XGBoost model, with its hyperparameter optimization, consistently outperformed both Random Forest and SVM in terms of accuracy, precision, recall, and F1 score. While Random Forest showed competitive results, especially in handling imbalanced data, it did not achieve the same level of optimization as XGBoost. The SVM, despite being effective in certain binary classification tasks, lagged behind XGBoost in performance for this dataset due to its sensitivity to the scaling of input features and hyperparameter tuning. All in all, **XGBoost** was the best choice for predicting kidney disease risk. To make sure we were using the most useful features, **Variance Threshold** technique was used to remove unnecessary data. We also used **SHAP (SHapley Additive exPlanations)** values to figure out which features had the biggest impact on our model's predictions. Some of the key factors that showed up as important were **albumin**, **hemoglobin**, **blood glucose**, and **age**, which match what we already know about kidney disease. According to the following chart; Albumin, hemoglobin, blood glucose, and age were some highly contributing factors to the development of kidney diseases. (Francis et al., 2024)

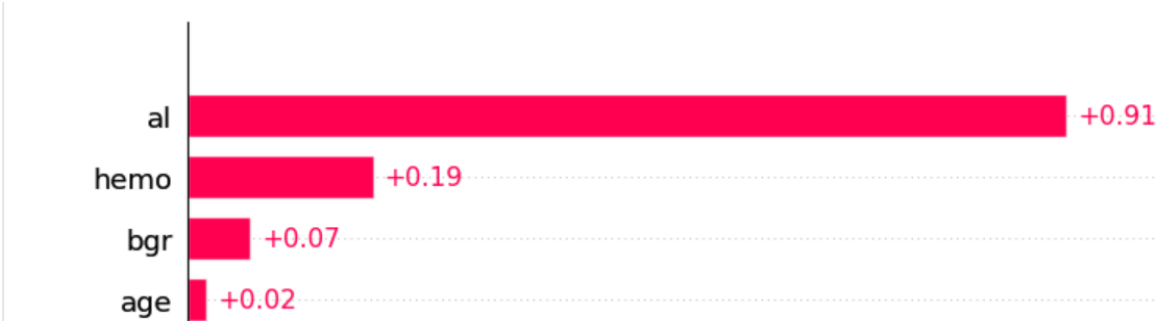


Figure 3.1: Snapshot of results of the SHAP Plot Demonstrating the Factors that Contributed the Most to the Development of Chronic Kidney Disease With 0 Being Not Affected and 1 Being Affected.

The SHAP analysis revealed that features such as blood glucose and age played a critical role in the model's predictions. Specifically, the higher the blood glucose level, the higher the likelihood of chronic kidney disease, highlighting the importance of monitoring diabetes as a risk factor. Age was another key predictor, with individuals over the age of 60 being at a significantly higher risk of developing kidney disease. These insights from SHAP values provide valuable understanding of the risk factors, further reinforcing the relevance of monitoring these health indicators in the prevention and early detection of kidney disease (Ahmed et al., 2022). Because of these factors, we can interpret which pre-existing conditions lead to a higher risk of kidney disease. For example, high levels of blood glucose are an indicator of diabetes, which causes 47% of new kidney disease cases. Additionally, Chronic Kidney

Disease typically occurs in people ages 60 or older. The level of albumin is important, as lower levels of the blood plasma are a possible result of kidney disease. The final factor that significantly contributed to the risk of chronic kidney disease is the levels of hemoglobin. Lower levels of hemoglobin can lead to anemia, which is a possible complication of kidney disease (Francis et al., 2024).

Overall **XGBoost** proved to be a powerful tool for predicting kidney disease risk. This study shows that **XGBoost** could be a game-changer for predicting kidney disease. While it's not perfect, it outperforms traditional methods and could help doctors detect kidney disease early, allowing for better treatment before it becomes severe. It offers a practical, cost-effective way to improve patient care and prevent the serious consequences of late-stage kidney disease.

6. Conclusion

To conclude, this research suggests, adding more types of data, like genetic information and environmental factors would help make the model more accurate and robust. To make the kidney disease prediction model even better, incorporating more diverse datasets like gathering data from different populations groups, different regions, socioeconomic backgrounds, and medical histories would be able to give more personalized predictions. Incorporating diverse datasets, such as data from varying geographical regions, socioeconomic statuses, and medical histories, would enhance the model's ability to account for a wide range of variables, providing more precise and personalized risk predictions. With the data inclusive of all these aspects, a holistic understanding of the risk factors and their interaction could be taken into account for more precise predictions (Friedman, 2019; Zhang, Fang, & Tran, 2023). These additional factors would allow the model to provide a deeper understanding of the risk factors, which would improve its ability to predict kidney disease risk at a greater accuracy for personalized treatments.

Another way to improve the model is by fine-tuning it even more and testing out different machine learning techniques and with vast amount of diverse data. Combining multiple models in an ensemble approach could help increase accuracy because it leverages the strengths of different algorithms, making the predictions more reliable. The goal is to turn the model into a user-friendly tool that people can use to input their health data and get personalized advice. This tool could help profile the patient and offer dietary tips like reducing sodium intake or suggest exercise routines based on individual risk factors. This would put people to make healthier choices and possibly prevent the need for expensive treatments in the future. Ultimately, this could help reduce healthcare costs by allowing for earlier interventions, all from the comfort of a person's own home. For this to be widely used, there needs to be an easy-to-use interface for both healthcare professionals and patients. This would allow doctors to input data, track patient health, and provide actionable recommendations. A dashboard for healthcare providers would be helpful, too, allowing them to interpret the data in real-time, making it a valuable tool in clinical settings.

Most importantly, the model should be able to learn and update itself with new data regularly. This ensures that it stays relevant and accurate as new trends, research, and methods emerge. A continuous learning system would allow the model to adapt to changes in health trends, and incorporating feedback from both patients and healthcare providers would help it improve over time. This feedback loop would help keep the model effective, ensuring that it keeps improving and providing real-world value.

Acknowledgments*:*

I would like to thank Dr. S. Prabhakar Karthikeyan, Professor, PDF (CPRI), Senior Member (IEEE), for his invaluable guidance and support throughout the course of this research. His expertise in research

publications greatly enhanced the depth and quality of this study.

7.Reference

1. Ahmed, S., Mothi, S. S., Sequist, T., Tangri, N., Khinkar, R. M., & Mendu, M. L. (2022). The Kidney Failure Risk Equation Score and CKD Care Delivery Measures: A Cross-sectional Study. *Kidney Medicine, 4*(1). https://doi.org/10.1016/j.xkme.2021.08.010
2. ESRD Treatment Choices (ETC) Model | CMS. (n.d.). *Centers for Medicare & Medicaid Services*. https://www.cms.gov/priorities/innovation/innovation-models/esrd-treatment-choices-model
3. Kidney Care Choices (KCC) Model | CMS. (n.d.). *Centers for Medicare & Medicaid Services*. https://www.cms.gov/priorities/innovation/innovation-models/kidney-care-choices-kcc-model
4. Francis, A., Harhay, M. N., Ong, A. C. M., Tummalapalli, S. L., Ortiz, A., Fogo, A. B., Fliser, D., Roy-Chaudhury, P., Fontana, M., Nangaku, M., Wanner, C., Malik, C., Hradsky, A., Adu, D., Bavanandan, S., Cusumano, A., Sola, L., Ulasi, I., & Jha, V. (2024). Chronic kidney disease and the global public health agenda: an international consensus. *Nature Reviews Nephrology*, 1–13. https://doi.org/10.1038/s41581-024-00820-6
5. Friedman, D. J. (2019). Genes and environment in chronic kidney disease hotspots. *Current Opinion in Nephrology and Hypertension, 28*(1), 87–96. https://doi.org/10.1097/MNH.0000000000000470
6. Gregorich, M., Kammer, M., Heinzel, A., Böger, C. A., Eckardt, K., Lambers Heerspink, H., Jung, B., Mayer, G., Meiselbach, H., Schmid, M., Schultheiß, U. T., Heinze, G., & Oberbauer, R. (2023). Development and Validation of a Prediction Model for Future Estimated Glomerular Filtration Rate in People With Type 2 Diabetes and Chronic Kidney Disease. *JAMA Network Open, 6*(4), e231870. https://doi.org/10.1001/jamanetworkopen.2023.1870
7. Harasemiw, O., Drummond, N., Singer, A., Bello, A., Komenda, P., Rigatto, C., Lerner, J., Sparkes, D., Ferguson, T. W., & Tangri, N. (2019). Integrating Risk-Based Care for Patients With Chronic Kidney Disease in the Community: Study Protocol for a Cluster Randomized Trial. *Canadian Journal of Kidney Health and Disease, 6*, 205435811984161. https://doi.org/10.1177/2054358119841611
8. New Model Predicts Cardiovascular Risk Among CKD Patients. (n.d.). *Penn Medicine*. Retrieved August 1, 2023, from https://www.pennmedicine.org/news/news-releases/2023/july/new-model-predicts-cardiovascular-risk-among-ckd-patients
9. Population Health and Chronic Kidney Disease: Improving Outcomes with Technology (2023). *Harvard Public Health Review, 45*, 67–80. https://www.hsph.harvard.edu/public-health-review/2023/08/population-health-and-chronic-kidney-disease-improving-outcomes-with-technology/
10. Research on Environment and Kidney Disease. (2023). *National Institute of Environmental Health Sciences*. https://www.niehs.nih.gov/health/topics/conditions/kidney-disease/index.cfm
11. Saran, R., Robinson, B., Abbott, K. C., Bragg-Gresham, J., Balkrishnan, R., Dietrich, X., & Pisoni, R. (2022). US Renal Data System 2022 Annual Data Report: Epidemiology of Kidney Disease in the United States. *American Journal of Kidney Diseases, 79*(4), A8–A18. https://doi.org/10.1053/j.ajkd.2022.01.012
12. Social Determinants of Health and Kidney Disease. (2023). *Kidney.org*. https://www.kidney.org/atoz/content/social-determinants-health-and-chronic-kidney-disease
13. Tangri, N., Grams, M. E., & Inker, L. A. (2020). Risk Stratification in Chronic Kidney Disease: Challenges and Opportunities. *Nephrology Dialysis Transplantation, 35*(2), 219–225. https://doi.org/10.1093/ndt/gfz153

14. Unger, T., Borghi, C., Charchar, F. J., Khan, N. A., Poulter, N. R., Prabhakaran, D., Schlaich, M. P., Stergiou, G. S., Tomaszewski, M., Touyz, R. M., Whelton, P. K., & Williams, B. (2020). 2020 International Society of Hypertension Global Hypertension Practice Guidelines. *Hypertension, 75*(6), 1334–1422. https://doi.org/10.1161/hypertensionaha.120.15026
15. US Centers for Medicare and Medicaid Services. (2024). Innovative Care Models to Improve Outcomes in CKD. *Centers for Medicare & Medicaid Services*. https://www.cms.gov/outreach-and-education/medicare-learning-network/mlnproducts/innovative-care-models-ckd
16. Zhang, Y., Fang, C., & Tran, M. (2023). Social and Environmental Predictors of Chronic Kidney Disease in Urban Settings. *Environmental Research, 224*, 115027. https://doi.org/10.1016/j.envres.2023.115027
17. Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z.-H., Naicker, S., Plattner, B., & Yang, C.-W. (2013). Chronic kidney disease: global dimension and perspectives. *The Lancet, 382*(9888), 260–272. https://doi.org/10.1016/S0140-6736(13)60687-X
18. Luyckx, V. A., & Brenner, B. M. (2005). The impact of the changing epidemiology of chronic kidney disease on public health. *Nature Clinical Practice Nephrology, 1*(7), 3–10. https://doi.org/10.1038/ncpcardio0394
19. Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E., & Hsu, C.-Y. (2014). Chronic kidney disease and the risk of death, cardiovascular events, and hospitalization. *The New England Journal of Medicine, 351*(13), 1296–1305. https://doi.org/10.1056/NEJMoa041031
20. Xie, Y., Yang, S., & Lou, J. (2021). Global, regional, and national burden of chronic kidney disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 395*(10225), 709–717. https://doi.org/10.1016/S0140-6736(20)32514-5
21. Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. *The Lancet, 379*(9811), 165–180. https://doi.org/10.1016/S0140-6736(11)60178-5
22. United Nations, Department of Economic and Social Affairs, Population Division. (2019). *World Population Prospects 2019: Highlights*. United Nations. https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf
23. Centers for Disease Control and Prevention (CDC). (n.d.). *CKD National Facts*. https://www.cdc.gov/kidney-disease/php/data-research/?CDC_AAref_Val=https://www.cdc.gov/kidneydisease/publications-resources/CKD-national-facts.html.

Doswell Health Informatics Conference

AI in Healthcare: Bridging Innovation, Practice and Education

April 25, 2025 | 8:30 a.m. - 4:30 p.m.

Conference Check-in | 7:30 - 8:30 a.m. | 1st Floor Lobby

Continental Breakfast Included

NOTE: TWU Dallas parking garage (https://map.concept3d.com/?id=576#!m/578263) floors 2-4 are reserved for conference attendees.

Agenda Overview

- **8:30 - 9 a.m.** | Welcome Remarks (https://twu.edu/doswell/agenda/#welcome)
- **9 - 10:15 a.m.** | Keynote Address, Judy Murphy. DN(hon), RN, FACMI, LFHIMSS, FAAN (https://twu.edu/doswell/agenda/#keynote)
- **10:30 a.m. - 12 p.m.** | Sessions A, B, C, D
  - Session A: AI in Health Informatics Education (https://twu.edu/doswell/agenda/#a)
  - Session B: Clinical Informatics and System Optimization (https://twu.edu/doswell/agenda/#b)
  - Session C: Mobile Health and Telehealth (https://twu.edu/doswell/agenda/#c)
  - Session D: Public Health Informatics and Social Determinants of Health (https://twu.edu/doswell/agenda/#d)
- **12 - 1 p.m.** | Lunch & Drawing (https://twu.edu/doswell/agenda/#lunch)
- **1 - 1:30 p.m.** | Poster Presentations and Exhibition Visits (https://twu.edu/doswell/agenda/#posters)
- **1:30 - 3 p.m.** | Sessions E, F, G, H
  - Session E: Data Science and Artificial Intelligence (https://twu.edu/doswell/agenda/#e)
  - Session F: Innovative Strategies in Health Education (https://twu.edu/doswell/agenda/#f)
  - Session G: AI in Clinical Practice and Care Delivery (https://twu.edu/doswell/agenda/#g)
  - Session H: Patient Safety and Workforce Issues (https://twu.edu/doswell/agenda/#h)

- **3:15 - 4 p.m.** | AI Tools (https://twu.edu/doswell/agenda/#ai)
- **4 - 4:30 p.m.** | Closing Remarks (https://twu.edu/doswell/agenda/#closing)

Welcome Remarks

Doswell Foundation Appreciation | 8:30 - 9 a.m. | Room 1010 (Auditorium)

Mikyoung A. Lee, PhD, RN

Professor, Doswell Endowed Chair for Informatics and Healthcare Transformation

Texas Woman's University Dallas, Texas

Monica Christopher

President, TWU Dallas

Texas Woman's University Dallas, Texas

Holly Hansen-Thomas, PhD

Vice Provost for Research, Innovation, & Corporate Engagement

Texas Woman's University Denton, Texas

Stephanie Woods, PhD, RN

Dean, College of Nursing

Texas Woman's University Denton, Texas

Keynote Address

9 - 10:15 a.m. | Room 1010 (Auditorium)

Judy Murphy, DN(hon), RN, FACMI, LFHIMSS, FAAN
Nurse Executive & Health IT Leader | Minneapolis, Minn.
Presentation Title: Health IT Strategy and Culture in the Era of AI

Judy Murphy will explore highs and lows of AI integration in healthcare, emphasizing both the IT strategy and organizational culture components of implementation. She will focus on the use of AI for patient outcome and experience improvements, as well as for workforce implications to optimize workflows, reduce administrative burden, and alleviate clinician burnout. Murphy will also examine issues surrounding the use of AI in healthcare, including: safety, ethics, transparency, governance, legislation/regulation, and staff competence.

More on Judy Murphy +

Murphy is a pioneer and thought leader in national and international health and nursing informatics communities. As Chief Nursing Officer for IBM Global Healthcare and Deputy National Coordinator for Programs and Policy in the Office of the National Coordinator for Health IT (ONC) in the U.S. Department of Health & Human Services, Murphy led efforts to assist health care providers in adopting health information technology and to promote consumers' greater understanding and use of health information technology for their own health.

Previously, she spent 36 years at Aurora Health Care in Wisconsin, 26 of those years in clinical informatics. As Vice President-Applications, she led the electronic health record program since 1995, when Aurora was an early adopter of health IT.

She has published over 100 articles and book chapters, with countless presentations nationally and internationally to her name. Murphy received an Honorary Doctor of Nursing degree from the University of Wisconsin-Milwaukee in 2022 and served on the AMIA and HIMSS Board of Directors. She is a Fellow in the American Academy of Nursing and the American College of Medical Informatics, and a HIMSS Lifetime Fellow.

The AMIA 2020 Virginia K. Saba Informatics Award, the HIMSS 2018 Most Influential Women in Health IT Award, the AMIA 2014 Don Eugene Detmer Award for Health Policy Contributions in Informatics, the HIMSS 2014 Federal Health IT Leadership Award, and the HIMSS 2006 Nursing Informatics Leadership Award are just a few awards that have recognized Murphy.

Session A

AI in Health Informatics Education

10:30 a.m. - 12 p.m. | Room 2102

Moderator: Susan Fenton, PhD, RHIA, ACHIP, FAMIA

Chance Reaves, MSN-Ed, RN

Senior Clinical Informatics Learning Specialist

Parkland Health Dallas, Texas

Heather DeGrande, PhD, CCRN-K

Associate Professor

Texas A&M University-Corpus Christi College of Nursing and Health Sciences Corpus Christi, Texas

Cora Rabe, DNP, CRNA

Program Director

University of Texas Medical Branch Galveston School of Nursing Galveston, Texas

Texas Woman's University College of Nursing Dallas, Texas

Stephanie Large, PhD, ANP-C, CNE

Assistant Professor

University of North Texas Health Science Center College of Nursing Fort Worth, Texas

Session B

Clinical Informatics and System Optimization

10:30 a.m. - 12 p.m. | Room 2702

Moderator: Joni Padden, DNP, RN

Ashley P. Huynh

Senior Business Analyst

UT Southwestern Medical Center Dallas, Texas

Myrna Garcia, MSN, RN, NI-BC, CPHQ

Informatics/Quality Improvement Specialist

Texas Children's Hospital Houston, Texas

Pratikshya Aryal, ND(BNYS)

MSHI Student

UT Southwestern Medical Center Dallas, Texas

Nelly Estefanie Garduno-Rapp, MD, MSHI

MSHI Program Associate Director, IMAGINe Lab Researcher

UT Southwestern Medical Center Dallas, Texas

Session C

Mobile Health and Telehealth

10:30 a.m. - 12 p.m. | Room 2706

Moderator: Sheila Haley, PhD, CFN

Latarsha S. Cheatham, DNP, APRN, FNP-BC

Assistant Dean of Graduate Studies, Assistant Professor

UTHealth Houston Cizik School of Nursing Houston, Texas

Texas Woman's University College of Nursing Houston, Texas

Sohi Kwon, PhD, RN, APN

Professor

Kyungpook National University Research Institute of Nursing Innovation Daegu, South Korea

Chanam Shin, PhD, RN

Associate Professor

Texas Woman's University College of Nursing Denton, Texas

Lauri Hix, MPH, RN

Graduate Assistant

University of Texas at Arlington College of Nursing and Health Innovation Arlington, Texas

Session D

Public Health Informatics and Social Determinants of Health

10:30 a.m. - 12 p.m. | Room 2707

Moderator: Theresa Mendoza

Sharlisa Raley, DNP, RN, NPD-BC

Nursing Professional Development Specialist

Methodist Health System Dallas, Texas

Shuhong Luo, EdD, MBA, RN

Chair, Department of Nursing Education & Associate Professor

Texas A&M University-Corpus Christi College of Nursing and Health Science Corpus Christi, Texas

Victor Kolade, MD

Associate in Internal Medicine

The Guthrie Clinic Sayre, Penn.

Lunch & Drawing | 12 - 1 p.m. | IHSD 1st Floor Lobby, 3rd Floor Cafeteria (Room 3620)

Poster Presentations and Exhibition Visits | 1 - 1:30 p.m. | IHSD 1st Floor Lobby

Poster Presenters and Exhibitors

Ruchira Garg
MSHI Student
UT Southwestern Medical Center
Dallas, Texas

Optimizing Pre-Treatment Dental Referrals for Head and Neck Cancer Patients

Mahesh Raisinghani, PhD, MBA
Professor
Texas Woman's University Merrilee Alexander Kick College of Business & Entrepreneurship
Denton, Texas

IT Transforming Healthcare: Trust, Security, and Privacy Issues Affected by AI among Patients, Employees, and Governance

Raaghul Subramani, MD
MS Data Science Student
University of Pittsburgh
Pittsburgh, Penn.

Predicting Duration of Hypotension for Early Sepsis Intervention

Yingzi Zhang, PhD, RN
Nurse Scientist
UT Southwestern Medical Center
Dallas, Texas

Natural Language Processing to Extract Acute Symptom Clusters from Triage Phone Notes with Cancer Patients

Junga Kim, PhD, RN

Kyungpook National University Hospital
Daegu, South Korea

Development of a Chronic Kidney Disease Prediction Model Using Machine Learning

Trish Jackson, APRN, WHNP-BC, EBP-C

DNP Student, Women's Health Nurse Practitioner
Texas Woman's University College of Nursing
Dallas, Texas

Utilizing Health Informatics to Optimize Access to Care in the Women's Health Ambulatory Setting: A Quality Improvement Project

Saiful Islam Badhon

PhD Student
University of North Texas, Department of Information Science
Denton, Texas

Empowering Rural Nurse Practitioners with a Large Language Model-Based Support System

Lakshmi Chandrabhanu

MSHI Student
UT Southwestern Medical Center
Dallas, Texas

Advancing Pediatric Suicide Prevention: A Clinician's Tool for Early Intervention

Ronald Samuel, MSN, RN

Nurse Manager
Texas Health Resources Hospital
Plano, Texas

The Impact of Artificial Intelligence to Determine Clinical Decisions and Pathways

Lori Camperlengo, RN, NI-BC, CHSE, CPHIMS

Clinical Assistant Professor
University of Texas at Arlington College of Nursing and Health Innovation
Arlington, Texas

Targeting Skills to Lead Academic EHR Implementation for Student Nurse Competency

Soumya Jayaraj

Biomedical Informaticist
UTHealth Houston, McWilliams School of Biomedical Informatics
Houston, Texas

Consolidated Clinical Document Architecture Quick Reference Sheet

Saiful Islam Badhon

PhD Student
University of North Texas, Department of Information Science
Denton, Texas

Acute Kidney Injury Anticipation using in ICU patients using Hybrid LSTM-Transformer

Kimberly Burt, RN

MSHI Student
UT Southwestern Medical Center
Dallas, Texas

Enhancing Nursing Unit and Blood Bank Communication for Blood Product Preparation Through WellSky and Vocera Integration

Aditi Dhariya, MSHI

Research Staff
UT Southwestern Medical Center
Dallas, Texas

Investigating Intrinsic Barriers to Clinical Practice Guidelines: An Examination of Ambiguity, Inconsistency, and Undecidability

Akshitha Gopikrishnan
High School Student
Rock Hill High School
Frisco, Texas

Assessing Kidney Disease Risk Using Environmental and Genetic Factors

Session E

Data Science and Artificial Intelligence

1:30 p.m. - 3 p.m. | Room 2102

Moderator: Ling Chu, MD

Mahesh Raisinghani, PhD, MBA, MSc

Professor

Texas Woman's University Merrilee Alexander Kick College of Business & Entrepreneurship
Denton, Texas

Yuhan Zhou

PhD Student

University of North Texas Department of Information Science Denton, Texas

Melika Rostami

PhD Student

University of North Texas Anuradha and Vikas Sinha Department of Data Science Denton, Texas

Harsha Snitha Kamineni

MSHI Student, Clinical Research Assistant

Indiana University Indianapolis, Ind.

Session F

Innovative Strategies in Health Education

1:30 p.m. - 3 p.m. | Room 2702

Moderator: Jennifer Witten, DNP, RN, NI-BC

Laura Kubin, PhD, RN, CPN, CHES, CNE, CHSE

Professor

Texas Woman's University College of Nursing Dallas, Texas

Hripsime Mantecon

PhD Student

University of North Texas Department of Information Science Denton, Texas

Meagan Rogers, PhD, RN, NPD-BC

Associate Professor

University of Texas at Arlington College of Nursing and Health Innovation Arlington, Texas

Jennifer Weber, MSN, RN

PhD Student

University of Texas at Arlington College of Nursing and Health Innovation Arlington, Texas

Session G

AI in Clinical Practice and Care Delivery

1:30 p.m. - 3 p.m. | Room 2706

Moderator: Linda Denke, PhD, RN, CCRC

Joni Padden, DNP, APRN

Chief Nursing Informatics Officer

Texas Health Resources Dallas, Texas

Kristin A. Raggio, RN, NI-BC

Manager of Clinical Informatics

Texas Tech University Health Science Center Lubbock, Texas

Baylor Scott & White Health System Dallas, Texas

Linda Denke, PhD, RN, CCRC

Nurse Scientist

UT Southwestern Medical Center Dallas, Texas

Session H

Patient Safety and Workforce Issues

1:30 p.m. - 3 p.m. | Room 2106

Moderator: Sharon Blackerby, DNP, RN, NPD-BC, NI-BC, CPHQ

Dawn-Maia Simmons, RN, CEN

Assistant Nurse Manager

Parkland Health and Hospital System - Hospital at Home Dallas, Texas

Joy Huang, MBA, RN, NEA-BC, CPHIMS

Regional Director of Clinical Informatics

Baylor Scott and White Health System Dallas, Texas

Morgan Bronson, CCLS

PhD Student

Texas Woman's University School of Human Sciences Denton, Texas

Cook Children's Medical Center Fort Worth, Texas

AI Tools

3:15 p.m. - 4 p.m. | Room 1010 (Auditorium)

Craig Limoli

CEO and Co-Founder

Wellsheet Newark, N.J.

Krish Purushothaman

Vice President & Practice Head of Implementation Services

CitiusTech Healthcare Worldwide

Closing Remarks, Evaluation and Drawing | 4 - 4:30 p.m. | Room 1010 (Auditorium)

Conference Planning Committee

Mikyoung A. Lee, PhD, RN

Planning Committee Chair

Professor, Doswell Endowed Chair for Informatics and Healthcare Transformation
Texas Woman's University College of Nursing
Dallas, Texas

Marion J. Ball, EdD, FACMI, FAAN, FIAHSI, FAHIMA, FMLA, FLHIMSS, FCHIME, FIMIA

Raj and Indra Nooyi Endowed Distinguished Chair in Bioengineering, Presidential Distinguished Professor and Executive Director
University of Texas at Arlington Center for Innovation in Health Informatics
Arlington, Texas

Eva Bernal, MBA

Nursing Operations Administrator
Texas Woman's University College of Nursing
Dallas, Texas

Sharon Blackerby, DNP, RN, NPD-BC, NI-BC, CPHQ

Clinical Assistant Professor & Professional Development Director
University of Texas at Arlington College of Nursing and Health Innovation
Arlington, Texas

Lorrie Burkhalter, MPH

Program Manager
UT Southwestern Medical Center

Dallas, Texas

Ling Chu MD

Associate Chief Medical Informatics Officer & Associate Professor, General Internal Medicine
UT Southwestern Medical Center
Dallas, Texas

Linda Denke, PhD, RN, CCRC

Nurse Scientist, Magnet Program & Nursing Research
UT Southwestern Medical Center
Dallas, Texas

Susan H. Fenton, PhD, RHIA, ACHIP, FAMIA

Dr. Doris L. Ross Professor, Vice Dean for Education, UT System Distinguished Teaching Professor
UTHealth Houston, McWilliams School of Biomedical Informatics
Houston, Texas

Joshua Flanagan

Digital Content Manager
Texas Woman's University
Denton, Texas

Ashley Hunsucker, RN

Clinical Solutions Specialist - Connexall
Adjunct Professor - Texas Woman's University College of Nursing
Dallas, Texas

Theresa Mendoza

Principal of Data Services & Research
DFWHC Education and Research Foundation
Dallas, Texas

Joni Padden, DNP, RN

Chief Nursing Information Officer
Texas Health Resources
Arlington, Texas

Ayesha Rahman, RN

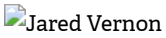
Ayesha Rahman, MSN, APRN, FNP-C

PhD Student & Research Assistant
Texas Woman's University College of Nursing
Denton, Texas


Ranjitha Rao, MBA

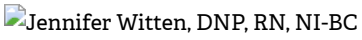
Ranjitha Rao, MBA

Budget/Financial Analyst
Texas Woman's University College of Nursing
Denton, Texas


Jared Vernon

Jared Vernon

Manager, TWU Dallas Campus & Technology
Texas Woman's University
Dallas, Texas


Jennifer Witten, DNP, RN, NI-BC

Jennifer Witten, DNP, RN, NI-BC

Division Director Clinical Informatics, HCA NTX Division
Medical City Healthcare
Dallas, Texas

Assessing Kidney Disease Risk Using Environmental and Genetic Factors

Akshitha Gopikrishnan, Rock Hill High School, Frisco, TX; Christie Martin, PhD, MPH, RN-BC, LHIT-HP, School of Nursing, University of Minnesota, Minneapolis, MN

Background/Introduction
Chronic Kidney Disease (CKD) is a disease that affects approximately 700 million people worldwide and has both environmental and genetic determinants. The genetic determinants include family history, gender, and age, while environmental determinants such as diet, toxins, and access to health care also play significant roles. Treatment has the best outcomes with early identification, but CKD typically remains undiagnosed until late in the disease, resulting in excessive morbidity and healthcare costs. This project aims to develop a machine-learning algorithm that integrates genetic and environmental factors for the prediction of CKD risk.

Purpose
The purpose of this project is to create a predictive algorithm that estimates the percent risk for CKD in an individual based on their genetic predispositions and environmental factors, allowing for early diagnosis and personalized intervention.

Methods
The Kaggle CKD dataset, comprising demographic information, medical history, and biomarkers, was used. Data preprocessing consisted of imputation of missing values and encoding categorical variables. The XGBoost classifier was used, and hyperparameter optimization was completed via GridSearchCV. Feature importance was calculated based on SHAP values. Certain vital biomarkers such as albumin levels, hemoglobin levels, blood glucose, and age were analyzed for their predictive capabilities in CKD risk prediction.

Results
XGBoost model had an accuracy of 1.0000, demonstrating outstanding predictive ability. SHAP analysis identified albumin level, hemoglobin level, blood glucose, and age as the most predictive features of CKD risk. Elevated blood glucose, decreased albumin and hemoglobin levels, and advanced age were related to high CKD risk.

Conclusions/Implications for Bioinformatics
This work demonstrates the potential of machine learning for the integration of genetic and environmental data for CKD risk prediction. The model provides a way of early detection with tailored health advice. Future work will include the addition of genetic sequencing data and additional environmental factors to the model, along with releasing it for general healthcare application and developing early intervention measures.

Reference

- Ahmed, S., Mothi, S. S., Sequist, T., Tangri, N., Khinkar, R. M., & Mendu, M. L. (2022). The Kidney Failure Risk Equation Score and CKD Care Delivery Measures: A Cross-sectional Study. *Kidney Medicine*, *4*(1). https://doi.org/10.1016/j.xkme.2021.08.010
- *ESRD Treatment Choices (ETC) Model | CMS*. (n.d.). Www.cms.gov. https://www.cms.gov/priorities/innovation/innovation-models/esrd-treatment-choices-model
- Francis, A., Harhay, M. N., Ong, A. C. M., Tummalapalli, S. L., Ortiz, A., Fogo, A. B., Fliser, D., Roy-Chaudhury, P., Fontana, M., Nangaku, M., Wanner, C., Malik, C., Hradsky, A., Adu, D., Bavanandan, S., Cusumano, A., Sola, L., Ulasi, I., & Jha, V. (2024). Chronic kidney disease and the

global public health agenda: an international consensus. *Nature Reviews Nephrology*, 1–13. https://doi.org/10.1038/s41581-024-00820-6

- Friedman, D. J. (2019). Genes and environment in chronic kidney disease hotspots. *Current Opinion in Nephrology and Hypertension*, *28*(1), 87–96. https://doi.org/10.1097/MNH.0000000000000470
- Gregorich, M., Kammer, M., Heinzel, A., Böger, C. A., Eckardt, K., Hiddo Lambers Heerspink, Jung, B., Mayer, G., Meiselbach, H., Schmid, M., Schultheiß, U. T., Heinze, G., & Rainer Oberbauer. (2023). Development and Validation of a Prediction Model for Future Estimated Glomerular Filtration Rate in People With Type 2 Diabetes and Chronic Kidney Disease. *JAMA Network Open*, *6*(4), e231870–e231870. https://doi.org/10.1001/jamanetworkopen.2023.1870
- Harasemiw, O., Drummond, N., Singer, A., Bello, A., Komenda, P., Rigatto, C., Lerner, J., Sparkes, D., Ferguson, T. W., & Tangri, N. (2019). Integrating Risk-Based Care for Patients With Chronic Kidney Disease in the Community: Study Protocol for a Cluster Randomized Trial. *Canadian Journal of Kidney Health and Disease*, *6*, 205435811984161. https://doi.org/10.1177/2054358119841611

Keywords
Chronic Kidney Disease, machine learning, risk prediction

MEDINFO 2025 — Healthcare Smart × Medicine Deep
M.S. Househ et al. (Eds.)

doi:10.3233/SHTI251145

Transforming Literature Reviews in Healthcare Research with Artificial Intelligence: Preliminary Insights

Christie L. MARTIN[a], Caitlin J. BAKKER[b], Jenna L. MARQUARD[a], Scott SITTIG[c], Jennifer ALLAIN[c], Akshitha D. GOPIKRISHNAN[d], Ming-Yuan CHIH[e], Kathy K. KIM[f], Vincent J. PETA[g], and Grace GAO[h]

[a] *University of Minnesota School of Nursing, Minneapolis, MN, USA,* [b] *University of Regina, Regina, SK, CANADA,* [c] *University of Louisiana at Lafayette, Lafayette, LA, USA,* [d] *Rock Hill High School, Frisco, TX, USA,* [e] *University of Kentucky, Lexington, KY, USA* [f] *Health Tequity LLC, Berkeley, CA, USA,* [g] *University of Minnesota Department of Surgery, Minneapolis, MN, USA,* [h] *Saint Catherine University, Saint Paul, MN, USA*

ORCiD IDs: Christie L. Martin https://orcid.org/0000-0003-1685-0205,
Caitlin J. Bakker https://orcid.org/0000-0003-4154-8382,
Jenna L. Marquard https://orcid.org/0000-0002-1256-6699,
Scott Sittig https://orcid.org/0000-0002-7179-152X,
Jennifer Allain https://orcid.org/0000-0001-5230-1030,
Akshitha D. Gopikrishnan https://orcid.org/0009-0007-7953-4919,
Ming-Yuan Chih https://orcid.org/0000-0001-7130-1415,
Kathy K. Kim https://orcid.org/0000-0001-5766-3938,
Vincent J. Peta https://orcid.org/0000-0002-9901-5207,
Grace Gao https://orcid.org/0000-0001-9198-1198

Abstract. A rapidly expanding array of Artificial Intelligence (AI) tools, with continually evolving features and functionalities, offers unprecedented opportunities to streamline literature reviews, expediting the screening, extraction, and synthesis phases. We present preliminary findings of evaluating various AI tools' strengths and limitations.

Keywords. Artificial intelligence, AI, literature review, evaluation

1. Introduction

A rapidly expanding array of Artificial Intelligence (AI) tools offers unprecedented opportunities in research and for innovative solutions to augment the literature review process. This poster highlights preliminary findings of the strengths and limitations of emerging AI tools and their capabilities to streamline complex review tasks.

2. Methods

We structured our evaluation of the AI tools based on 3 phases of the literature search and review process: screening, extraction, and evidence synthesis. We selected and assessed tools specifically designed to support the review based on ChatGPT-generated strengths and limitations.

3. Results/Discussion

Rayyan, [1] a web-based AI tool for the initial screening, features a user-friendly interface and team-based collaborative tools, including AI-generated relevance scores. While efficient, it may lead to overconfidence in automated relevance assessments. *NotebookLM*,[2] an AI-driven tool, is well-suited for extraction and synthesis. Key features include document support, AI summarization, and interactive querying. Its effectiveness depends on input quality. *Research Rabbit*, [3] an AI-based tool for literature discovery and network visualization, supports users in organizing and synthesizing data. It is most effective in resource gathering but may require supplemental tools. *ChatGPT*[4] has the potential to support many review phases by offering efficient organization, synthesis, and summarization. The breadth of functionality can be valuable, though the tool cannot appraise study quality and may misrepresent information. *Elicit*,[5] an AI tool for synthesis, aims to automate repetitive tasks, including screening, extraction, and synthesis. It is customizable but may be challenging for new users.

4. Conclusion

Each AI tool brings unique strengths to the literature review process, and their utility is best realized when combined with human oversight and expertise. It is essential to combine AI and human judgment when creating reliable, timely evidence.

References

[1] Rayyan Systems Inc. Rayyan [web application]. Rayyan Systems Inc.; 2025. Accessed March 21, 2025.
[2] Google LLC. NotebookLM [AI research assistant]. Google; 2025. Accessed March 21, 2025. https://notebooklm.google
[3] ResearchRabbit Inc. ResearchRabbit [research discovery tool]. ResearchRabbit Inc.; 2025. Accessed March 21, 2025. https://www.researchrabbit.ai
[4] OpenAI. ChatGPT [Large language model]. OpenAI; 2025. Accessed March 21, 2025. https://chat.openai.com.
[5] Ought. Elicit [The AI Research Assistant]. Ought; 2025. Accessed March 21, 2025. https://elicit.com.

Achieving Value-based Care Success with an AI Driven SDOH Actionability Score

Kerry Son, MSN, BSN; **Rivka Atadja**, DNP, RN; **Matt Wilson**, DPT; **Derek Olsson**; **Anant Vasudevan**, MD, MBA; Radial, Concord, MA

Although the gap in outcomes between those with and without social risk factors has narrowed, disparities persist. For entities in value based care arrangements, CMS has put forth guidelines for the documentation and use of SDOH data, making it more important than ever. Despite this, existing data sources and tools have many shortcomings when it comes to social risk adjustment. Programs that link payment to outcomes, such as those established by the Affordable Care Act, have led to improvements in many quality measures. However, these improvements have not been uniform across all patient groups. Health systems and payers need a mechanism to consistently and accurately identify and adjust for social risk as well as a simple, synthesized assessment of how to mitigate the impacts for any given patient.

Alliance for Nursing Informatics (ANI) Emerging Leader Project: Enhancing Nurse Competency and Advocacy in the Age of Artificial Intelligence

Ann M. Wieben, PhD, RN, NI-BC, University of Wisconsin, Madison, WI, USA

The increasing integration of AI tools into healthcare settings necessitates active nursing involvement. AI, as defined by the World Health Organization, is “a branch of computer science, statistics and engineering that uses algorithms or models to perform tasks and exhibit behaviors such as learning, making decisions and making predictions.” While AI offers potential benefits in healthcare, such as improved diagnostics and workflow optimization, it also presents complex challenges related to safety, ethics, and equity. The nursing profession needs to be actively involved in shaping how AI tools are evaluated, used, and governed in nurse practice and education settings. Nurses must possess AI literacy and competencies to shape the future of AI in healthcare. AI literacy, as defined by Long and Magerko, involves “a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace” with competencies being understood as the understanding, knowledge or skills needed to perform a task.

Artificial Intelligence Tools in Healthcare

Carolyn S. Harmon, PhD, DNP, RN, NI-BC, University of South Carolina, Columbia, SC, USA
Heather Carter-Templeton, PhD, RN, NI-BC, FAAN, West Virginia University, Morgantown, WV
Leyla Pordeli, DNP, MBA, MSN, NI-BC, RN Jacksonville University, Jacksonville, FL, USA
Tonya Judson, DNP, RN, NI-BC, CNE, University of Alabama at Birmingham, Birmingham, AL, USA
Amy M. Rosa, DNP, MSMI, RN, Sentara, Virginia Beach, Virginia, USA
Katherine Taylor Pearson, DNP, RN, NI-BC, CPHIMS, CLSSBB, CPBI, CKM, Texas Tech University Health Sciences Center, School of Nursing, Lubbock, Texas, USA

Artificial Intelligence (AI) is redefining healthcare, with tools ranging from diagnostic imaging systems to predictive analytics platforms. These technologies promise to improve clinical efficiency, patient outcomes, and healthcare accessibility. However, AI's potential to exacerbate health disparities underscores the need for frameworks that emphasize equity, transparency, and ethical responsibility. Drawing from recent systematic reviews, there is evidence to suggest AI tools are being implemented across diverse healthcare settings. Large language tools such as ChatGPT and Co-pilot have shown promise in simplifying patient communication, offering tailored health education, and streamlining administrative workflows. IBM Watson Health has been instrumental in oncology care, providing treatment recommendations aligned with clinical

guidelines and peer-reviewed evidence. Case studies, including AI-based breast cancer screening tools, highlight their enhanced sensitivity and ability to detect malignancies in underserved populations. The analysis integrates frameworks like the FUTURE-AI Medical AI Algorithm Checklist to evaluate tool performance, adaptability, and bias mitigation.

Bridging the Gap:
The Need for Nursing Informatics Programs in Ugandan Public Universities

Aidah Nanvuma, BNS, PGDPPM, MPH, University of Minnesota
Charles Osingada, MA, MPH, PhD, University of Minnesota
Rebecca Wurtz, MD, MPH, IDI; University of Minnesota

Nursing informatics is a fundamental discipline that leverages the nursing body of knowledge, information technology, and management to advance patient care outcomes, simplify health care administration, and ease decision-making in the clinical setting 1. Although the field of nursing informatics is steadily growing globally, its adoption and implementation in Africa remain challenged by the limitations in infrastructure, training, low investments, and weak policy frameworks that guide the processes. A significant knowledge gap in nursing informatics in Africa is the limited understanding of how nursing informatics content is integrated into pre- and post-licensure nursing education programs and how it is taught within these programs. This leads to an inadequately prepared workforce to meet the increasing informatics demands in clinical settings. There is limited research on how nursing informatics is incorporated into pre- and post-licensure nursing programs in Uganda and its impact on preparing nurses for real-world clinical settings. This study aims to investigate integrating nursing informatics education into pre- and post-licensure nursing programs in Uganda.

Clustering Fibromyalgia Patients Using the All of Us Research Program

Marisa Sileo, MSN, RN, NI-BC, Georgia Institute of Technology, Atlanta, GA
Victoria Menzies, PhD, APRN, FAAN, University of Florida, Gainesville, FL
Steven G. Johnson, PhD, University of Minnesota, Minneapolis, MN
Lisiane Pruinelli, PhD, RN, FAMIA, University of Florida, Gainesville, FL

Fibromyalgia is a chronic pain disorder that significantly impacts quality of life. With no known cure, treatment is focused on managing symptoms. Despite fibromyalgia’s impact on pain and quality of life, it remains challenging to diagnose and manage effectively. This study sought to characterize fibromyalgia patient clusters, identifying factors and patterns in the fibromyalgia population and their relationship with patient-reported outcomes.

Collecting and Evaluating Perioperative Physiological Waveforms from the Patients during Major Surgeries

Dan Li, Ph.D, RN (Corresponding author)
Fei Zhang, Ph.D, CRNA, RN, School of Nursing, University of Pittsburgh, PA, USA

There has been an increase in anesthesia-related mortality among older adults with morbidities undergoing surgery. Anesthesia devices measure physiological waveforms, but it is challenging to acquire and store this data for clinical research due to limited interoperability and integration with AIMS. AIMS can only record low-fidelity physiological data for medical records and billing, but high-fidelity data like waveforms is often not captured and saved.

Navigating the Digital Tightrope: Nurses Perspectives on Fall Prevention Technology

Quetina R. Jones, MSN-ED, BSN, RN, CNEn, University of Florida, College of Nursing
Nicole M. Cope, MEd., EdS., LMHC, BSN, University of Florida, College of Nursing
Joslyn Sofia Massie, BSc, BSN, University of Florida, College of Nursing
Kimberly Martinez, MSN, RN, CCRN, UF Health Shands Hospital
Ragnhildur I. Bjarnadottir, PhD, MPH, RN, University of Florida, College of Nursing

Falls in the acute care setting continually present a major threat to patient safety. Technology has become increasingly integrated into healthcare, offering valuable tools for fall prevention. However, the implementation of these technologies, and how it impacts nursing practice and clinical decision-making in fall prevention, have not been well explored. This study aimed to describe nurses' perspectives on the current state of fall prevention in an acute care hospital. This abstract focuses on a subset of findings from the study related to nurses' perspectives on the benefits and challenges of technology use in fall prevention.

Normal Mixtures Analysis Reveals Racial and Ethnic Differences in Hospital Breastfeeding Support and Breastfeeding Intensity

Tiffany T. Gallagher, BA, IBCLC, PhD(c), University of Minnesota School of Nursing
Robin R. Austin, PhD, DNP, RN-BC, FAMIA, University of Minnesota School of Nursing
Carrie E. Neerland, PhD, APRN, CNM, University of Minnesota School of Nursing
Michelle A. Mathiason, MS, University of Minnesota School of Nursing
Ellen W. Demerath, PhD, University of Minnesota School of Public Health
Wendy S. Looman, PhD, APRN, CPNP-PC, University of Minnesota School of Nursing
Anne C. McKechnie, PhD, RN, University of Minnesota School of Nursing

Inequitable care may contribute to persistent disparities in breastfeeding rates across different racial and ethnic groups in the United States.1 The purpose of this study was to analyze the differences in breastfeeding support provided and feeding outcomes of infants using electronic health records (EHR) from nine urban and rural hospitals within a single hospital system in Minnesota.

Nurses and mHealth App Development

Akshitha D. Gopikrishnan, Rock Hill High School, Frisco, TX
Caitlin J. Bakker, MLIS, AHIP-D, Dr. John Archer Library & Archives, University of Regina, Canada
Melissa S. Breth, DNP, Breth Consulting, L.L.C., Aurora, IL
Whende M. Carroll, MSN, RN, NI-BC, FHIMSS, Healthcare Information and Management Systems Society, Informatics/Government Relations, Chicago, IL
Grace Gao, PhD, DNP, NI-BC, LHIT-HP, School of Nursing, St. Catherine University, St Paul, MN
Mikyoung A. Lee, PhD, RN, College of Nursing, Texas Woman's University, Dallas, TX
Victoria L. Tiase, PhD, RN, University of Utah School of Medicine, Department of Biomedical Informatics, Salt Lake City, UT
Tami H. Wyatt, PhD, RN, ANEF, FAAN, College of Nursing, University of Tennessee, Knoxville, TN
Christie L. Martin, PhD, MPH, RN-BC, LHIT-HP, University of Minnesota School of Nursing

The prevalence of mobile health (mHealth) applications (apps) continues to increase, with an estimated 350,000 apps worldwide. Although apps have the potential to improve patient care, nurses' involvement in app development remains minimal. This exclusion leads to the development of apps that fail to engage both nurses and patients as end users, resulting in apps that often overlook patient-centered care principles and fail to adequately address clinical needs.

Review of Standards-based Electronic Case Reporting (eCR) for Public Health Surveillance

Chanhee Kim, RN, MPH, CIC, University of Minnesota, Minneapolis
Larry Chen, BS, Tufts University School of Medicine, Boston, Massachusetts
Sarah Solarz, MPH, Minnesota Department of Health, Saint Paul, Minnesota
Sripriya Rajamani, MBBS, PhD, MPH, FAMIA, University of Minnesota, Minneapolis

Efficient surveillance of notifiable infectious diseases is critical for public health, yet traditional reporting methods (phone/fax/paper) are delayed, incomplete and inefficient. Electronic Case Reporting (eCR), built on HL7 standards, automates case reporting from healthcare to public health, enhancing timeliness, accuracy, and scalability of disease surveillance. This study evaluates the impact of eCR on key metrics: timeliness, completeness, and reporting volume.

The Scope of Nursing Informatics in Africa: A Systematic Review

Aidah Nanvuma, BNS, PGDPPM, MPH (c) Infectious Diseases Institute, Makerere University, University of Minnesota
Richard Kwizera, PhD, Infectious Diseases Institute
Charles Osingada, MA, MPH, PhD, University of Minnesota
Rebecca Wurtz, MD, MPH, University of Minnesota

Nursing informatics is crucial in modern healthcare, and it integrates information science and technology to enhance nursing education, clinical decision-making, and patient outcomes. However, in Africa, the adoption and implementation of nursing informatics remain limited due to infrastructural, educational, and systemic challenges. A lack of formal training, insufficient resources, and inadequate policies impede progress. Understanding these barriers and identifying potential solutions is essential for advancing nursing informatics in the region.

A Theoretical Data Science Framework for Neonatal Care: Advancing Predictive Analytics in the NICU

Janet Northcote, RN, PhD Student, University of Florida, Gainesville, FL, USA
Lisiane Pruinelli, PhD, RN, MS, FAMIA, University of Florida, Gainesville, FL, USA

The NICU is a high-acuity environment where clinicians must make critical, time-sensitive decisions while managing vast amounts of patient data. Despite the availability of electronic health records (EHRs), real-time physiologic monitoring, and clinical documentation, neonatal decision-making remains predominantly reactive rather than proactive. This is due to fragmented data systems, unpredictable patient trajectories, and limited integration of predictive decision-support tools. Care inefficiencies not only increase the cognitive workload of NICU nurses but also contribute to delayed recognition of clinical deterioration, inconsistencies in care delivery, and challenges in resource allocation. Additionally, neonatal critical care imposes significant emotional and financial burdens on families and healthcare systems. The proposed adapted theory presents a structured, scalable data science framework that leverages predictive analytics to enable early risk detection, reduce cognitive burden, and optimize nursing decision-making in the NICU.

The Unique Nurse Identifier (UNI)

Nancy J Beale, PhD, RN, NI-BC, FAMIA, Catholic Health Services of Long Island, Melville, New York, USA

There has been debate about which number should be used as a unique nurse identifier, which has led to confusion and division among professional stakeholders. In April 2024, a task force comprised of representatives from the American Nurses Association (ANA), the Alliance for Nursing Informatics (ANI), the National Council of State Boards of Nursing (NCSBN), and the Nursing Knowledge Big Data Science Policy and Advocacy Workgroup (NKBDS) identified a collaborative solution for use as a Unique Nurse Identifier that comprised both the NCSBN ID and the NPI.

Using Machine Learning to Identify Developmental Screening, Delay Diagnosis, and Service Use Patterns

Megan A. Foradori, PhD, RN, VA Quality Scholar, Interprofessional Improvement Research Education Evaluation Clinical Center; VA Northeast Ohio Healthcare System, Cleveland, Ohio; Case Western Reserve University, School of Nursing, Cleveland, Ohio, USA
Valerie Boebel Toly, PhD, RN, CPNP-PC, FAAN, Case Western Reserve University, School of Nursing, Cleveland, Ohio
Faye A. Gary, EdD, RN, FAAN, Case Western Reserve University, School of Nursing, Cleveland
Barbara A. Lewis, PhD, CCC-SLP, Case Western Reserve University, Department of Communication Sciences, Cleveland, Ohio
Nicholas K. Schiltz, PhD, Case Western Reserve University, School of Nursing, Cleveland, Ohio

One in six children in the United States have developmental disabilities, with signs of delay often appearing in infancy or toddlerhood. Despite modest gains in screening rates, less than 20% of children with known delays receive early intervention or special education services, and those remaining are left to struggle with undiagnosed and untreated deficits until they are identified upon entering kindergarten. Machine learning can help identify these children by uncovering patterns in large datasets that smaller studies might miss.

Nurses and mHealth App Development

Akshitha D. Gopikrishnan, Rock Hill High School, Frisco, TX
Caitlin J. Bakker, MLIS, AHIP-D, Dr. John Archer Library & Archives, University of Regina, Regina, SK, Canada
Melissa S. Breth, DNP, Breth Consulting, L.L.C., Aurora, IL
Whende M. Carroll, MSN, RN, NI-BC, FHIMSS, Healthcare Information and Management Systems Society, Informatics/Government Relations, Chicago, IL
Grace Gao, PhD, DNP, NI-BC, LHIT-HP, School of Nursing, St. Catherine University, St Paul, MN
Mikyoung A. Lee, PhD, RN, College of Nursing, Texas Woman's University, Dallas, TX
Victoria L. Tiase, PhD, RN, University of Utah School of Medicine, Department of Biomedical Informatics, Salt Lake City, UT
Tami H. Wyatt, PhD, RN, ANEF, FAAN, College of Nursing, University of Tennessee, Knoxville, TN
Christie L. Martin, PhD, MPH, RN-BC, LHIT-HP, School of Nursing, University of Minnesota Minneapolis, MN

Nurses and mHealth App Development

Learning Objectives:

- Appraise the role of nurses in the Software Development Life Cycle.
- Assess strategies for integrating nursing expertise throughout app development.
- Identify gaps in interdisciplinary collaboration for mHealth innovation.

Introduction: The prevalence of mobile health (mHealth) applications (apps) continues to increase, with an estimated 350,000 apps worldwide. Although apps have the potential to improve patient care, nurses' involvement in app development remains minimal. This exclusion leads to the development of apps that fail to engage both nurses and patients as end users, resulting in apps that often overlook patient-centered care principles and fail to adequately address clinical needs.

Methods: Our research was guided by the Software Development Life Cycle (SDL), which includes 7 phases of app development: planning, requirements analysis, design, coding, testing, deployment, and maintenance. To understand the current state of nurse participation in mHealth innovation, specifically app development, we used data from a previously published literature review,[1] a grey literature search, and surveys of nurses, app developers, and professionals in the mHealth app industry.

Results: Our study revealed that only 13% of surveyed app companies reported nurse participation in any development phase. Nurse participation was typically limited to 3 phases: requirements gathering, design, and testing. Despite their expertise, nurses are not extensively engaged across the entire SDL, particularly in critical phases such as coding and deployment. A detailed review of published and online resources confirmed that nurses primarily serve as subject matter experts or evaluators. Barriers to nurse participation included a lack of interdisciplinary collaboration, insufficient training in technical tools, and limited organizational prioritization of nurses' roles in innovation.[2]

Discussion and Conclusion: The findings underscore the failure to capitalize on nurses' specialized expertise—patient advocacy, health literacy, and person-centered care—in key phases of mHealth app development. Addressing this gap would improve the clinical relevance, safety, and nurse and patient adoption of apps. To achieve this, it is critical to include nurse-led initiatives in mHealth innovation, fostering collaborative opportunities for nurses to interact with designers and developers and encouraging health organizations to prioritize nurses' contributions during all phases of development. Recognizing the essential role of nurses in digital innovation paves the way for systemic change, resulting in tools that enhance patient outcomes, build trust, and advance the future of digital healthcare.

References

1. Bakker CJ, Wyatt TH, Breth MC, Gao G, Janeway LM, Lee MA, Martin CL, Tiase VL. Nurses' Roles in mHealth App Development: Scoping Review. JMIR Nurs. 2023 Oct 17;6:e46058. DOI: 10.2196/46058 PubMed Central PMCID: PMC10618897.

2. Wyatt TH, Bakker CJ, Breth MS, Carroll WM, Gao G, Lee MA, Martin CL, Tiase VL. Nurses and mHealth App Development [White Paper]. Creative Commons License. 2024. Available from: https://hdl.handle.net/11274/16615

11/16/2025 | *8:30 AM – 12:00 PM* | Room 7

W29: Leveraging Artificial Intelligence for Evidence Synthesis: Screening, Extraction, and Synthesis

Presentation Type: Workshop

Leveraging Artificial Intelligence for Evidence Synthesis: Screening, Extraction, and Synthesis

Presentation Time: 08:30 AM - 11:30 AM

Abstract Keywords: Artificial Intelligence, Information Retrieval, Information Extraction
Working Group: Consumer Health Informatics Working Group
Primary Track: Foundations

This three-hour workshop explores the integration of Artificial Intelligence (AI) into the evidence synthesis process,
including systematic reviews and meta-analyses, essential for quickly evolving informatics and digital health research. AI has the potential to enhance efficiency by automating key tasks, but researchers must carefully assess
AI's limitations, such as inaccuracies, biases, or "hallucinations," to ensure the integrity of synthesized evidence.
This workshop will provide a structured overview of the evidence synthesis process, with participants introduced to
both traditional workflows and AI-based tools for different phases of evidence synthesis. Through hands-on activities, participants will work in small groups to apply AI tools in evidence synthesis while evaluating the advantages, challenges, and ethical considerations of these tools.

The workshop is structured around three key phases, where participants will learn to use, test, and evaluate the pros and cons of various AI tools:

1. Literature Search and Screening: systematically searching databases and screening studies based on predefined criteria to identify relevant literature.
2. Data Extraction: collecting key information from selected studies, such as study design, population, intervention, and outcomes.
3. Data Synthesis: analyzing and integrating extracted data to identify patterns, relationships, and gaps in the evidence base, forming the basis for conclusions.

The workshop will conclude with a group discussion and Q&A session, allowing participants to reflect on their experience, discuss ethical considerations, and critically evaluate AI-generated outputs. By the end of the workshop,
participants will have a comprehensive understanding of how AI can support evidence synthesis and will gain practical insights into responsible, transparent, and effective AI use.

Speakers:
Christie Martin, PhD, MPH, RN-BC, LHIT-HP
University of Minnesota School of Nursing

Grace Gao, PhD, DNP
St Catherine University

Jenna Marquard, PhD
University of Minnesota

Scott Sittig, PhD, MHI, RHIA

Leveraging Artificial Intelligence for Evidence Synthesis: Screening, Extraction, and Synthesis
Workshop - Collaborative

Date: Sunday (11/16)
Time: 8:30 AM to 12:00 PM
Room: Room 7

Back to Speaker Gallery

University of Louisiana at Lafayette

Vishala Mishra, MBBS
Duke

Julie Doberne, MD, PhD
Oregon Health & Science University

Zainab Balogun, MS, MA
University of Maryland Baltimore County

Authors:
Christie Martin, PhD, MPH, RN-BC, LHIT-HP - *University of Minnesota School of Nursing*; **Grace Gao, PhD, DNP** - *St Catherine University*; **Jenna Marquard, PhD** - *University of Minnesota*; **Scott Sittig, PhD, MHI, RHIA** - *University of Louisiana at Lafayette*; **Vishala Mishra, MBBS** - *Duke*; **Julie Doberne, MD, PhD** - *Oregon Health & Science University*; **Zainab Balogun, MS, MA** - *University of Maryland Baltimore County*; **Sayantani Sarkar, PhD** - *Yale University*; **Ming-tse Tsai, MD MSHI** - *Kura Care*; **Akshitha Gopikrishnan, High School Student** - *Rock Hill High School*; **Caitlin Bakker, PhD(c), MLIS, AHIP-D** - *Dr. John Archer Library, University of Regina*;

Section Editor: Susan Alexander, DNP, ANP-BC, ADM-BC

Innovative Technologies
Optimizing Nursing Practice

Victoria L. Tiase, PhD, RN, Christie L. Martin, PhD, MPH, RN-BC, LHIT-HP, Caitlin J. Bakker, MLIS, AHIP-D, Suzanne S. Fink, MSN, RN, Grace Gao, PhD, DNP, Akshitha D. Gopikrishnan, Mikyoung A. Lee, PhD, RN, Clair Lunt, DHSc, Katherine Taylor-Pearson, DNP, RN, Tami H. Wyatt, PhD, RN, Whende M. Carroll, MSN, RN

Healthcare in the United States is rapidly evolving due to increasingly digitized healthcare settings.[1] Digital health tools—computing platforms, connectivity, software, and sensors—are being developed and procured by health systems to create efficiencies in healthcare processes while lowering the cost of care.[2] Digital health technologies are changing how nurses practice, from robots delivering medications to voice-controlled patient rooms.[3] The increased availability of such tools presents an opportunity to innovate in a way that benefits patient care and supports clinicians at the point of care (POC).

Although the responsibility is shared, nurses are considered the primary health information technology (HIT) users. Nurses are well-qualified to inform developers because their direct interaction with patients allows them to observe inefficiencies, identify potential issues, and prioritize patients' needs.[4] By connecting technological advancements and innovative solutions to healthcare needs, nurses can uncover clinical barriers that limit the full potential of HIT.[5]

Three pervasive HIT advancements that have arisen in the last decade include artificial intelligence (AI), monitoring sensors (hereafter, sensors), and virtual reality (VR). At the POC, these digital companions improve clinical decision-making, increase patient monitoring, and augment nursing interventions. Using real-world examples and forecasting uses of HIT tools, we aim to provide insights and guidance on how the nursing profession can innovate patient-centric digital solutions.

Author Affiliations: School of Medicine, University of Utah, Salt Lake City (Dr Tiase); Population Health and Systems Cooperative, University of Minnesota School of Nursing (Martin); Dr. John Archer Library & Archives, University of Regina (Bakker); School of Nursing and Health Professions, Cuyahoga Community College (Fink); School of Nursing, St. Catherine University (Dr Gao); Rock Hill High School (Gopikrishnan); College of Nursing, Texas Woman's University (Dr Lee); Digital and Technology Department, Mount Sinai Health System (Lunt); Texas Tech University Health Sciences Center (Dr Taylor-Pearson); College of Nursing, University of Tennessee (Dr Wyatt); Healthcare Information and Management Systems Society, Chicago, IL (Carroll).

V.L.T. and C.L.M. are co-first authors.

V.L.T. is currently receiving funding from NINR, project #1K01NR021256-01. The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.

Corresponding author: Victoria L. Tiase, PhD, RN, School of Medicine, University of Utah, 421 Wakara Way, Salt Lake City, UT 84111 (victoria.tiase@utah.edu).

DOI: 10.1097/CIN.0000000000001352

KEY POINTS

- Because of their skills and roles in patient care, nurses are well-positioned to use and envision future capabilities of emerging technologies.
- Nurses currently leverage health information technology tools such as artificial intelligence, monitoring sensors, and virtual reality.
- Nurses' adaptability and central role in digital health initiatives are crucial for enhancing patient outcomes and integrating technological advancements into healthcare practice.

ARTIFICIAL INTELLIGENCE AND CLINICAL DECISION SUPPORT SYSTEMS

Artificial Intelligence

With recent technological advancements, AI has become more accessible and capable, expanding the role of HIT in clinical nursing practice. This has sparked significant debate about its use in healthcare and the necessary safeguards.[6] AI supports personalized, evidence-based, and efficient nursing care.[7] Although the rise of generative AI has brought AI into the mainstream, we will take a broad view of AI, explore its applications in clinical decision support, and discuss how generative AI can create efficiencies for nurses.

At its core, AI is the science of developing intelligent computer programs.[8] It enables machines to learn and perform tasks that typically require human decision-making.[9,10] AI can be broadly categorized into traditional AI and generative AI. Traditional AI leverages historical data to make predictions, identify patterns, and automate tasks through predefined algorithms. In contrast, generative AI, such as OpenAI's ChatGPT and DALL-E, creates and summarizes text, voice dialogue, and other forms of content. These two approaches are not mutually exclusive and may be used synergistically to enhance nursing practice. For example, traditional AI can support clinical decision-making by analyzing patient data, whereas generative AI can assist with nursing documentation, patient education materials, and personalized communication. Despite the limited availability of clinical decision support systems (CDSSs) designed specifically