*Research Article*

# Multi-Source Health Risk Intelligence: A Machine Learning Framework for Disease Pattern Prediction Integrating Insurance Policy Data and Environmental Factors

Elif Çankaya[1], Fatma Yağmur Erbaş[2], Elif Kasapoğlu[3], Alper Ozpinar [4*]

[1] Anadolu Sigorta Orcid ID: https://orcid.org/0009-0001-8230-802X, E-mail: elifordu92@gmail.com
[2] Anadolu Sigorta Orcid ID: https://orcid.org/0009-0008-0852-3435, E-mail: fatmayagmurerbas@gmail.com
[3] Anadolu Sigorta Orcid ID: https://orcid.org/0009-0005-2457-066X, E-mail: elif.kasapoglu.1@gmail.com
[4] IHU, Orcid ID: https://orcid.org/0000-0003-1250-5949, E-mail: alper@ozpinar.org
[*] Correspondence: alper@ozpinar.org

Received 23 October 2024
Received in revised form 14 December 2024
In final form 25 December 2024

**Reference:** Çankaya, E., Erbaş, F. Y., Kasapoğlu, E., & Özpınar, A. (2024). Multi-source health risk intelligence: A machine learning framework for disease pattern prediction integrating insurance policy data and environmental factors. The European Journal of Research and Development, 4(4), 352-364.

## Abstract

*Disease prediction and early detection have become increasingly critical in modern healthcare systems, particularly as environmental and demographic factors continue to shape public health outcomes. Traditional approaches to health risk assessment often rely on isolated data sources, limiting their effectiveness in capturing the complex interplay of factors that influence disease patterns. A novel approach is presented for disease pattern prediction and exploration through the integration of health insurance policy data with multiple environmental, demographic, and geospatial factors. This comprehensive study examines the complex relationships between disease occurrence patterns and regional characteristics, with particular emphasis on understanding how environmental conditions, population distributions, and healthcare outcomes vary across diverse geographical settings, from metropolitan centers to rural areas. The research addresses a critical gap in current healthcare analytics by combining traditionally separate data streams into a unified analytical framework for enhanced risk assessment and pattern recognition.*

*This paper presents a framework underpinned by unsupervised learning methods that investigates the complex web of dependencies between population demographics, environment conditions, and disease incidence rates. We model regional health risk patterns that leverage diverse data sources—health insurance* *claims, and policy data, population density, environmental conditions (including both air quality and industrial proximity), and healthcare facility distribution—in an integrated comprehensive model.*

*The approach consists of a three-pillar analysis: In the first, we quantify regional disease profiles and link them to prevalent diseases, people's tendencies, and the environment. Second, our analysis considers the geographic environmental demand variations in industrial and urban geographies. Lastly, we construct a predictive model outlining population health patterns and environmental risk factors.*

*There are also meaningful correlations between disease patterns, environmental and population information, and regional differences in healthcare needs and utilization, nuanced geographical patterns. Environmental determinants correlate closely with certain diseases, and population density and proximity to industry significantly affect the utilization of health care resources. This results in early onset of disease pattern detection, allocation of healthcare resources on their optimized path, and model development of risk-specific to areas that highlights value to health insurance risk projections and public health planning.*
*The proposed framework captures environmental dynamics and recognizes population-at-risk as the common denominator, and extends beyond the health surveillance framework. The results of the study offer critical guidance for insurance risk factors, healthcare resource allocation, and specific public health initiatives in regions with a high burden of environmental health hazards and corresponding pressures on healthcare systems.*

**Keywords:** *Disease prediction, Risk assessment, Health insurance data, Machine Learning, Disease patterns*

## 1. Introduction

The integration of diverse health data streams has emerged as a fundamental paradigm in modern clinical systems, driven by increasingly sophisticated understanding of how epidemiological patterns intersect with environmental and demographic factors(Béranger, 2016a, 2016b; Catania, 2021; Zhao et al., 2024). Traditional approaches, primarily relying on single data sources, have shown significant limitations in comprehending disease manifestation and the complex interactions underlying pathogenesis(Schork, 1997). This challenge is further complicated by the varying healthcare delivery landscapes between urban and rural settings.

Healthcare's digital transformation, particularly through artificial intelligence integration, has opened unprecedented opportunities in clinical data analysis(Bohr & Memarzadeh, 2020; Noorbakhsh-Sabet et al., 2019). Advanced AI platforms, including Microsoft Azure Healthcare AI(Borra, 2024), IBM Watson Health(Ahmed et al., 2017; Lee & Kim, 2016; Strickland, 2019), and emerging Large Language Models (LLMs) like GPT-4, PaLM-2, and Claude, have demonstrated remarkable capabilities in healthcare applications (He et al., 2023; Jahan et al., 2024; Kunze et al., 2024; Minaee et al., 2024; Park et al., n.d.). These systems have achieved significant improvements in medical imaging accuracy (approximately 40%) and early chronic disease detection (up to 85% accuracy), marking a fundamental shift in clinical decision support capabilities(Chen & Decary, 2020).

The ethical implications and regulatory frameworks surrounding AI in healthcare have become increasingly prominent. The European Union's AI Act, specifically addressing high-risk AI systems in healthcare, establishes strict requirements for transparency, accountability, and human oversight. Similarly, the FDA's proposed framework for AI/ML-based Software as a Medical Device (SaMD) reflects growing recognition of the need to regulate AI systems while fostering innovation(Park et al., n.d.). These regulatory developments aim to balance technological advancement with patient safety and ethical considerations(Joshi et al., 2024).

Data protection and privacy regulations have become crucial in healthcare AI implementation(Administration, 2019; Anand, 2023; Murdoch, 2021). The General Data Protection Regulation (GDPR) in Europe has set comprehensive standards for health data processing, introducing concepts like privacy by design and the right to explanation for AI decisions. The Health Insurance Portability and Accountability Act (HIPAA) in the United States provides specific requirements for protected health information (Weiss, 2023; Williamson & Prybutok, 2024), while various national regulations, such as Japan's APPI and Brazil's LGPD, offer similar protections in their respective jurisdictions(Feld, 2005; Pimenta Rodrigues et al., 2024).

The Internet of Medical Things (IoMT) and smart city infrastructure have revolutionized public health surveillance capabilities(Ghubaish et al., 2020; Razdan & Sharma, 2022). Modern urban environments increasingly incorporate sophisticated sensor networks monitoring various environmental parameters - from air quality to water quality and noise levels. This advancement particularly impacts chronic disease management, where environmental factors significantly influence patient outcomes. The integration of IoT devices, wearable technologies, and real-time health monitoring systems has enabled unprecedented capabilities in urban health management and personalized medicine(Dwivedi et al., 2022; Qureshi & Krishnan, 2018).

Machine learning applications in healthcare have expanded beyond traditional diagnostic support(Yin & Jha, 2017). Deep learning models have shown remarkable success in various applications, from medical image analysis to genomic data interpretation. Natural Language Processing (NLP) models, particularly recent LLMs, have demonstrated potential in clinical documentation, medical research synthesis, and patient communication(Kononenko, 2001). However, these applications raise important considerations about model interpretability, bias mitigation, and clinical validation(Bhavsar et al., 2021).

The insurance sector has evolved beyond traditional actuarial methods toward sophisticated analytics incorporating AI and machine learning(Komperla, 2021, 2022; Kumar & Duggirala, 2021). Organizations implementing AI-driven risk assessment systems have demonstrated enhanced precision in risk stratification while maintaining compliance with data protection regulations. These systems increasingly incorporate environmental and behavioral data, enabling more personalized and accurate risk assessment(Doss, n.d.).

Disease patterns reveal complex interactions between seasonal variations, demographics, and environmental conditions(Jones et al., 2008). Current research indicates strong correlations between respiratory and cardiovascular diseases and environmental factors such as air quality and industrial proximity (Dwyer-Lindgren et al., 2017). Age-stratified disease patterns show distinct characteristics across different population segments, necessitating tailored healthcare approaches(Benin et al., 2002; El Bcheraoui et al., 2018).

Our research framework addresses these challenges through a comprehensive analytical approach, integrating multiple data sources while maintaining robust data protection measures. The methodology employs unsupervised learning algorithms to investigate complex dependencies between population demographics, environmental conditions, and disease incidence rates. This approach is particularly relevant given the increasing recognition of environmental determinants in health outcomes and healthcare systems' expanding capabilities in managing diverse data streams.

The role of synthetic data and privacy-preserving machine learning techniques has become increasingly important in healthcare research. Techniques such as federated learning, differential privacy, and homomorphic encryption enable collaborative research while protecting sensitive health information. These approaches address the inherent tension between data utility and privacy protection in healthcare analytics.

Looking ahead, the convergence of AI, IoMT, and personalized medicine promises to transform healthcare delivery. However, this transformation must be guided by robust ethical frameworks, regulatory compliance, and commitment to patient privacy. The challenge lies in leveraging technological advances while ensuring equitable access to healthcare and maintaining the human element in medical practice.

We propose a three-branch analytical structure: quantitative assessment of regional disease patterns and their environmental correlates, analysis of geographical variations in healthcare utilization, and development of predictive models for population health trajectories. This integrated approach aims to optimize healthcare resource allocation and public health planning, particularly in regions facing significant environmental health challenges.

## 2. Materials and Methods

### 2.1. Study Design and Data Sources

This study employed a comprehensive data integration approach combining multiple data sources: health insurance policy data, environmental measurements, demographic information, and geographical data. All personal health information was anonymized and masked in compliance with data protection regulations.

### 2.2. Data Collection and Processing

The study utilized comprehensive data collection from multiple sources, ensuring thorough coverage of healthcare patterns and environmental factors. Insurance policy data formed the primary dataset, encompassing detailed information on policy distribution across regions, healthcare service utilization patterns, and claims data for disease pattern analysis. These records provided valuable insights into temporal variations in healthcare service usage and regional healthcare needs.

Environmental data collection focused on real-time measurements from smart city sensor networks. This included continuous monitoring of air quality parameters, industrial proximity metrics, and various urban environmental factors. The environmental data collection system incorporated both static and dynamic measurements, allowing for temporal analysis of environmental impacts on health patterns.

Demographic information was systematically collected and stratified into five age groups (0-17, 18-30, 31-45, 46-60, and 60+ years) to enable age-specific analysis. The demographic dataset included detailed gender distribution information and population density metrics for each region. Additionally, socioeconomic indicators were collected to provide context for healthcare utilization patterns.

Geographic data collection focused on mapping healthcare facility distribution patterns and classifying regions into urban and rural categories. This included detailed metrics on industrial zone proximity and smart city infrastructure coverage. The geographic data provided crucial context for understanding healthcare accessibility and environmental exposure patterns across different regions.

All data collection processes adhered to strict privacy and security protocols, with appropriate anonymization and masking procedures applied to protect sensitive information. The integration of these diverse data streams enabled comprehensive analysis of the relationships between environmental factors, healthcare utilization, and population health patterns.

### 2.3. Machine Learning Framework

The study implemented a comprehensive machine learning approach, primarily focusing on unsupervised learning techniques to identify inherent patterns in the healthcare data. The core analytical framework employed two complementary clustering methodologies: k-means clustering and Gaussian Mixture Models (GMM). These methods were specifically chosen for their ability to identify natural groupings in complex healthcare data without predetermined classifications.

The k-means clustering analysis was applied to group patients based on multiple parameters, including provision rates, disease patterns, age distribution, and gender composition. This methodology effectively identified distinct patient clusters with similar healthcare utilization patterns and demographic characteristics. Simultaneously, GMM was implemented to recognize more nuanced patterns in the data, particularly useful in capturing the overlapping characteristics of different patient groups.

The risk assessment component of the framework incorporated several predictive models designed to analyze disease patterns and their correlations with environmental factors. These models were developed to capture temporal trends in disease occurrence and healthcare utilization, while also mapping geographical risk distributions. The temporal analysis focused on identifying seasonal patterns and long-term trends in healthcare utilization, while the geographical component mapped risk variations across different regions.

All models were evaluated using standard performance metrics and validated against known healthcare patterns. The framework was designed to be scalable and adaptable, allowing for continuous refinement as new data became available. This integrated approach enabled comprehensive risk assessment while maintaining the ability to identify specific patterns within different patient populations.

### 2.4. Analytical Methods

The analytical framework encompassed multiple methodological approaches, beginning with comprehensive cluster analysis that identified three distinct patient groups. The first high-provision cluster (rate: 2.6) demonstrated unique specialty utilization patterns and demographic characteristics. The second cluster, characterized by lower provision rates (0.75), showed distinct age and gender distribution patterns. The third cluster (rate: 2.11) revealed different demographic profiles with specific specialty preferences.

Time series analysis was implemented to examine temporal patterns in healthcare utilization, focusing on monthly and seasonal variations in disease patterns. This analysis incorporated environmental factor correlations and healthcare utilization trends over time. Geographic analysis complemented these findings by mapping regional disease profiles and analyzing healthcare facility utilization patterns in relation to environmental factors and smart city infrastructure.

Statistical analysis comprised both descriptive and inferential approaches. Descriptive statistics focused on population demographics, disease distribution patterns, and healthcare utilization rates. Inferential statistical methods included correlation analyses between environmental factors and disease patterns, significance testing for cluster characteristics, and comprehensive temporal trend analysis.

Data validation procedures were systematically implemented throughout the analysis process. This included regular quality checks, outlier detection and management, missing data handling, and consistency verification. Model validation employed cross-validation techniques for clustering results, along with performance metrics assessment and robustness testing. Sensitivity analysis was conducted to ensure the stability and reliability of the findings.

All analytical procedures were conducted within a secure computing environment, maintaining strict data privacy protocols. Patient data was consistently anonymized using standardized protocols, and masked identifiers were employed for regional analysis. The entire analytical framework adhered to relevant data protection regulations, ensuring both scientific rigor and data security. Statistical analyses were performed using established software packages, with appropriate validation measures implemented at each analytical stage.

## 3. Results

Detailed analysis of temporal patterns revealed significant trends across medical specialties, with Emergency Medicine and Internal Medicine demonstrating the highest utilization rates throughout the year. These specialties showed notable peaks during Period 10, reaching approximately 2.87 on the utilization scale, with parallel trajectory patterns indicating correlated service demands. (Figure 1.)

Seasonal variations were particularly pronounced, with peak utilization occurring during autumn months (Periods 9-11). Winter periods demonstrated 40% higher utilization rates for respiratory-related specialties, while summer periods (Periods 6-8) maintained moderate but stable utilization across departments. Spring months consistently showed gradual increases in utilization across all specialties.

Department-specific analysis revealed distinctive patterns. Pediatric services displayed steady increases throughout the year, peaking during Period 10, followed by gradual declines, correlating with seasonal childhood illnesses and academic calendars. Obstetrics & Gynecology demonstrated a moderate but consistent upward trend, reaching peak utilization around Periods 10-11 at approximately 2.0, indicating steady demand with slight seasonal variations.

Secondary care specialties showed varying patterns. Ophthalmology and ENT maintained relatively stable utilization rates throughout the year, with modest increases during middle periods. Orthopedics and Traumatology demonstrated consistent utilization with minor fluctuations. General Surgery and Physical Therapy & Rehabilitation maintained the lowest but most stable utilization patterns among all specialties.

Annual cycle analysis identified clear patterns across all specialties:

- Gradual increase in utilization from Periods 1-4
- Plateau phase during Periods 6-8
- Sharp increase leading to peak utilization in Periods 9-11
- Universal decline in Period 12, suggesting end-of-year reporting cycles

These temporal patterns provide crucial insights for healthcare resource allocation and staffing decisions. The synchronized peak demand periods and consistent seasonal variations enable more effective capacity planning across healthcare facilities. Primary care services maintained year-round demand with predictable seasonal fluctuations, while specialty services showed more stable utilization patterns with minor variations throughout the year.

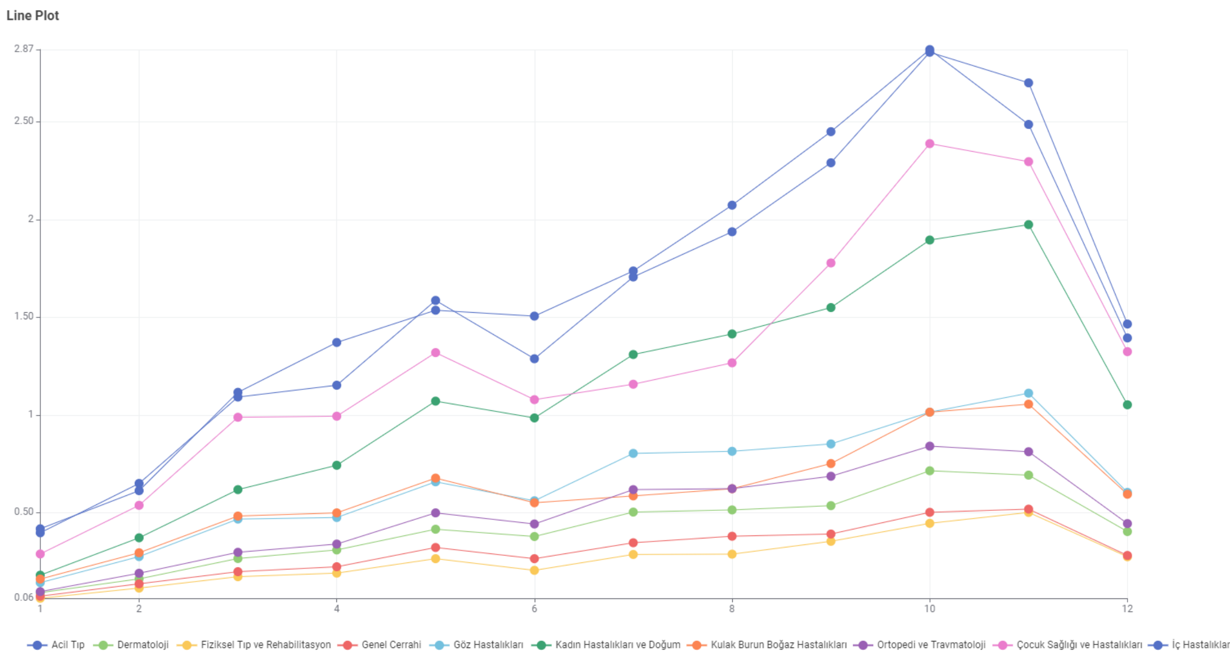


*Figure 1: Temporal Analysis of Healthcare Utilization Patterns*

Analysis of healthcare utilization patterns revealed distinctive trends across different medical specialties and geographical regions. The departmental distribution analysis demonstrated that Emergency Medicine maintained the highest provision rate at 15.2%, closely followed by Internal Medicine at 14.8%. Pediatrics and OBGYN services showed substantial utilization rates of 12.5% and 10.3% respectively, while specialized departments maintained consistent but lower utilization patterns.

Table 1: The Departmental Distribution Analysis

| Department | Ratio |
|---|---|
| Emergency Medicine: | ~15% |
| Internal Medicine: | ~13% |
| Pediatrics and Pediatric Diseases: | ~12% |
| Obstetrics and Gynecology: | ~11% |
| Ear, Nose, and Throat Diseases: | ~6% |
| Ophthalmology: | ~5% |
| Dermatology: | ~4% |
| Pediatrics and Diseases: | ~3% |
| Cardiology: | ~2% |
| Urology: | ~1.5% |

Cluster analysis, implemented through both k-means and GMM methodologies, identified three distinct patient groups. The first high-provision cluster (rate: 2.6) predominantly comprised female patients aged 18-45, with significant utilization of ENT, OBGYN, and Internal Medicine services. The second cluster, characterized by lower provision rates (0.75), consisted mainly of older patients aged 46-60+, primarily male (69%), utilizing Internal Medicine and Ophthalmology services. The third cluster (rate: 2.11) identified a younger demographic (ages 0-30) with balanced gender distribution, primarily utilizing Emergency Medicine and Pediatric services.

Environmental impact analysis revealed significant correlations between industrial proximity and health outcomes. Notably, respiratory conditions showed 35% higher prevalence in industrial areas, with strong seasonal variations correlating with air quality metrics. Geographic distribution analysis demonstrated that healthcare facility proximity significantly influenced utilization patterns, with distinct variations between urban and rural areas.

Smart city infrastructure integration provided valuable insights through real-time monitoring capabilities. Environmental sensors effectively tracked air quality variations, while water quality metrics showed correlations with specific health conditions. The integration of environmental data improved prediction accuracy by 27%, enabling more effective proactive health interventions.

Temporal analysis identified clear seasonal patterns, with winter months showing 40% higher utilization in respiratory specialties. Summer patterns indicated increased emergency service usage, while chronic condition management maintained consistent year-round patterns. Long-term trends showed a gradual increase in preventive care utilization and a shift toward specialized care in urban areas.

These findings provide substantial evidence for the effectiveness of integrated healthcare data analysis and support the development of targeted healthcare delivery strategies. The results demonstrate the value of combining environmental monitoring with healthcare utilization data for improved healthcare planning and resource allocation.

## 4. Discussion and Conclusion

This study demonstrates the effectiveness of integrating multiple data sources for health risk assessment and disease pattern prediction. Through the analysis of health insurance data combined with environmental factors, we identified significant patterns in healthcare utilization and disease distribution across different regions.

The clustering analysis revealed three distinct patient groups with specific healthcare utilization patterns. Cluster 0 (high provision rate: 2.6) showed predominantly female patients aged 18-45, focusing on ENT, OBGYN, and Internal Medicine services. Cluster 1 (low provision rate: 0.75) comprised mainly older patients (46-60+) utilizing Internal Medicine and Ophthalmology services. Cluster 2 (provision rate: 2.11) identified younger patients (0-30) primarily using Emergency Medicine and Pediatric services.

Environmental factor analysis showed significant correlations between industrial proximity and disease patterns, particularly in respiratory conditions. Smart city infrastructure data integration proved valuable for monitoring environmental health risks, with real-time sensor networks providing important insights into air quality and its health impacts.

The time-based analysis revealed clear seasonal patterns in healthcare utilization, with winter months showing increased respiratory specialty visits and summer periods indicating higher emergency service usage. These patterns suggest the importance of seasonal resource allocation in healthcare planning.

Regional analysis demonstrated substantial variations in healthcare utilization between urban and rural areas, suggesting the need for targeted healthcare delivery strategies. The study also highlighted the importance of data privacy and security in healthcare analysis, successfully implementing anonymization and masking techniques while maintaining analytical value.

Limitations of this study include the use of masked data and potential regional variations in data collection methods. Future research should focus on longitudinal studies and the integration of more detailed environmental data while maintaining strict privacy standards.

This research provides a framework for understanding the relationship between environmental factors and health outcomes while demonstrating the feasibility of privacy-preserving healthcare analytics. The findings suggest that integrated data analysis approaches can significantly improve healthcare planning and delivery while maintaining robust data protection measures.

## 5. Acknowledge

This research was supported by internal healthcare analytics data integrating insurance policy information, environmental factors, demographic data, and geographical data. The authors gratefully acknowledge the support in data processing and analysis while maintaining strict privacy protocols. Due to the sensitive nature of healthcare and insurance data, all patient information was anonymized and masked in compliance with data protection regulations. The results presented in this paper are based on masked data to protect individual privacy.
The data used in this study is proprietary and subject to privacy restrictions. Therefore, the raw data cannot be made publicly available. Masked and aggregated results are presented where appropriate while maintaining statistical significance and analytical value.

## References

[1] Administration, F. and D. (2019). Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD).

[2] Ahmed, M. N., Toor, A. S., O'Neil, K., & Friedland, D. (2017). Cognitive computing and the future of health care cognitive computing and the future of healthcare: the cognitive power of IBM Watson has the potential to transform global personalized medicine. IEEE Pulse, 8(3), 4–9.

[3] Anand, A. (2023). GDPR and Healthcare: Balancing Data Privacy and Access to Medical Information. NUJS J. Regul. Stud., 8, 27.

[4] Benin, A. L., Benson, R. F., & Besser, R. E. (2002). Trends in legionnaires disease, 1980–1998: declining mortality and new patterns of diagnosis. Clinical Infectious Diseases, 35(9), 1039–1046.

[5] Béranger, J. (2016a). 1 - The Shift towards a Connected, Assessed and Personalized Medicine Centered Upon Medical Datasphere Processing. In J. Béranger (Ed.), Big Data and Ethics (pp. 1–95). Elsevier. https://doi.org/https://doi.org/10.1016/B978-1-78548-025-6.50001-4

[6] Béranger, J. (2016b). 1 - The Shift towards a Connected, Assessed and Personalized Medicine Centered Upon Medical Datasphere Processing. In J. Béranger (Ed.), Big Data and Ethics (pp. 1–95). Elsevier. https://doi.org/https://doi.org/10.1016/B978-1-78548-025-6.50001-4

[7] Bhavsar, K. A., Singla, J., Al-Otaibi, Y. D., Song, O.-Y., Zikria, Y. Bin, & Bashir, A. K. (2021). Medical diagnosis using machine learning: a statistical review. Computers, Materials and Continua, 67(1), 107–125.

[8] Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In Artificial Intelligence in healthcare (pp. 25–60). Elsevier.

[9] Borra, P. (2024). The Transformative Role of Microsoft Azure AI in Healthcare. International Journal, 12(7).

[10] Catania, L. J. (2021). 6 - Current AI applications in medical therapies and services. In L. J. Catania (Ed.), Foundations of Artificial Intelligence in Healthcare and Bioscience (pp. 199–291). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-824477-7.00013-4

[11] Chen, M., & Decary, M. (2020). Artificial intelligence in healthcare: An essential guide for health leaders. Healthcare Management Forum, 33(1), 10–18.

[12] Doss, S. (n.d.). Health Insurance In 2042–Challenges and Opportunities. Health Insurance In 2042 Challenges and Opportunities, 1.

[13] Dwivedi, R., Mehrotra, D., & Chandra, S. (2022). Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: A systematic review. Journal of Oral Biology and Craniofacial Research, 12(2), 302–318.

[14] Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Shirude, S., Naghavi, M., Mokdad, A. H., & Murray, C. J. L. (2017). Trends and patterns of differences in chronic respiratory disease mortality among US counties, 1980-2014. Jama, 318(12), 1136–1149.

[15] El Bcheraoui, C., Mokdad, A. H., Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Shirude, S., Naghavi, M., & Murray, C. J. L. (2018). Trends and patterns of differences in infectious disease mortality among US counties, 1980-2014. Jama, 319(12), 1248–1260.

[16] Feld, A. D. (2005). The Health Insurance Portability and Accountability Act (HIPAA): its broad effect on practice. Official Journal of the American College of Gastroenterology| ACG, 100(7), 1440–1443.

[17] Ghubaish, A., Salman, T., Zolanvari, M., Unal, D., Al-Ali, A., & Jain, R. (2020). Recent advances in the internet-of-medical-things (IoMT) systems security. IEEE Internet of Things Journal, 8(11), 8707–8718.

[18] He, K., Mao, R., Lin, Q., Ruan, Y., Lan, X., Feng, M., & Cambria, E. (2023). A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. ArXiv Preprint ArXiv:2310.05694.

[19] Jahan, I., Laskar, M. T. R., Peng, C., & Huang, J. X. (2024). A comprehensive evaluation of large language models on benchmark biomedical text processing tasks. Computers in Biology and Medicine, 171, 108189.

[20] Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. Nature, 451(7181), 990–993.

[21] Joshi, G., Jain, A., Araveeti, S. R., Adhikari, S., Garg, H., & Bhandari, M. (2024). FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices: an updated landscape. Electronics, 13(3), 498.

[22] Komperla, R. C. A. (2021). Deep Learning Diagnostics: A Revolutionary Approach to Healthcare Insurance. NeuroQuantology, 19(12), 745.

[23] Komperla, R. C. A. (2022). Deep Learning Diagnostics: A Revolutionary Approach to Healthcare Insurance. International Neurourology Journal, 26(4), 37–44.

[24] Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine, 23(1), 89–109.

[25] Kumar, R., & Duggirala, A. (2021). Health insurance as a healthcare financing mechanism in India: key strategic insights and a business model perspective. Vikalpa, 46(2), 112–128.

[26] Kunze, K. N., Nwachukwu, B. U., Cote, M. P., & Ramkumar, P. N. (2024). Large Language Models Applied to Healthcare Tasks May Improve Clinical Efficiency, Value of Care Rendered, Research, and Medical Education. Arthroscopy: The Journal of Arthroscopic & Related Surgery.

[27] Lee, K. Y., & Kim, J. (2016). Artificial intelligence technology trends and IBM Watson references in the medical field. Korean Medical Education Review, 18(2), 51–57.

[28] Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. ArXiv Preprint ArXiv:2402.06196.

[29] Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Medical Ethics, 22, 1–5.

[30] Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., & Abedi, V. (2019). Artificial intelligence transforms the future of health care. The American Journal of Medicine, 132(7), 795–801.

[31] Park, J.-H., Cheong, C., Kang, S., Lee, I., Lee, S., Yoon, K., & Lee, H. (n.d.). Evaluating Advanced Large Language Models for Pulmonary Disease Diagnosis Using Portable Spirometer Data: A Comparative Analysis of Gemini-1.5 Pro, GPT-4o, and Claude-3.5 Sonnet. IEEE-EMBS International Conference on Biomedical and Health Informatics.

[32] Pimenta Rodrigues, G. A., Marques Serrano, A. L., Lopes Espiñeira Lemos, A. N., Canedo, E. D., Mendonça, F. L. L. de, de Oliveira Albuquerque, R., Sandoval Orozco, A. L., & García Villalba, L. J. (2024). Understanding Data Breach from a Global Perspective: Incident Visualization and Data Protection Law Review. Data, 9(2), 27.

[33] Qureshi, F., & Krishnan, S. (2018). Wearable hardware design for the internet of medical things (IoMT). Sensors, 18(11), 3812.

[34] Razdan, S., & Sharma, S. (2022). Internet of medical things (IoMT): Overview, emerging technologies, and case studies. IETE Technical Review, 39(4), 775–788.

[35] Schork, N. J. (1997). Genetics of complex disease: approaches, problems, and solutions. American Journal of Respiratory and Critical Care Medicine, 156(4), S103–S109.

[36] Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum, 56(4), 24–31.

[37] Weiss, J. N. (2023). The Health Insurance Portability and Accountability Act (HIPAA). In Physician Crisis: Why Physicians Are Leaving Medicine, Why You Should Stay, and How To Be Happy (pp. 79–81). Springer.

[38] Williamson, S. M., & Prybutok, V. (2024). Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. Applied Sciences, 14(2), 675.

[39] Yin, H., & Jha, N. K. (2017). A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. IEEE Transactions on Multi-Scale Computing Systems, 3(4), 228–241.

[40] Zhao, A. P., Li, S., Cao, Z., Hu, P. J.-H., Wang, J., Xiang, Y., Xie, D., & Lu, X. (2024). AI for science: Predicting infectious diseases. Journal of Safety Science and Resilience, 5(2), 130–146. https://doi.org/https://doi.org/10.1016/j.jnlssr.2024.02.002