DS Journal of Artificial Intelligence and Robotics
Volume 3 Issue 1, 9-17, Jan - Mar 2025
ISSN: 2583-9926 / https://doi.org/10.59232/AIR-V3I1P102

Original Article

An Expert System Based on Rules to Assist Physicians in the Screening and Early Differential Diagnosis of Cancer Patient

Humberto Cuteso Matumueni

Institute Superior University Nimi Ya Lukeni, Department of Computer Science Soyo, Angola.

hkuteso@hotmail.com

Received: 05 January 2025; Revised: 01 February 2025; Accepted: 03 March 2025; Published: 28 March 2025

Abstract -This article aims to help doctors diagnose patients suffering from early cancer, which is essential to improve patients' chances of survival and treatment options. However, early diagnosis remains complex due to the diversity of symptoms and medical examinations, which are often difficult to interpret. A rule-based expert system is important to help doctors diagnose a patient's cancer early. This system relies on a structured medical knowledge base and an inference engine that applies clinical rules to generate diagnostic recommendations according to the case. This article presents a rule-based expert system designed for the early diagnosis of breast, lung, and colorectal cancers. To create a knowledge base, patient data was collected from oncologists. After collecting the data, a rule base was developed in collaboration with oncologists and integrated into the system's inference engine. Finally, the system was implemented and validated in hospitals and clinics. Thus, the models were validated individually by comparing the results obtained with reality over a long period of time, and the rules were tested by apparent validation and compared with the decisions made by experts. The system's evaluation test result showed a high sensitivity of 59%. The system reduced the number of false positives and helped doctors identify high-risk patients more quickly. The rule-based expert system represents a major advancement in the medical field and could significantly improve early cancer detection.

Keywords - Knowledge-based system, Cancer, Artificial Intelligence, Expert system, Exsys Corvid.

1. Introduction

Cancer remains a leading cause of death globally, with approximately ten million deaths in 2020, or nearly one in six deaths. Early diagnosis is essential to improve survival rates, as it allows for more effective therapeutic management. However, despite technological and medical advances, early cancer diagnosis still presents significant challenges, both globally and in Africa. Globally, several barriers hinder early cancer diagnosis. These include issues such as lack of awareness and knowledge; many people are unaware of the early signs and symptoms of cancer, which delays medical consultation; limited access to health services; in many regions, especially in low- and middle-income countries, access to diagnostic services is restricted due to the lack of infrastructure and trained personnel and the high cost of diagnostic tests. Screening and diagnostic procedures can be expensive, limiting their accessibility for a large part of the population.

In Africa, these challenges are exacerbated by additional factors such as the shortage of specialized healthcare professionals the continent suffers from a notable shortage of oncologists and medical personnel trained in cancer diagnosis and treatment, inadequate medical infrastructure, many healthcare facilities lack essential equipment for early cancer screening and diagnosis, sociocultural factors, cultural beliefs and lack of awareness often lead to underestimation of symptoms and delay in seeking medical care and the absence of reliable cancer registries, the lack of accurate data on cancer incidence and prevalence hinders the planning and implementation of effective disease control strategies. All these obstacles contribute to late diagnosis and high cancer-related mortality rates in Africa. Therefore, developing innovative solutions adapted to the African context is imperative to facilitate early diagnosis.Cancer remains a major cause of morbidity and mortality worldwide despite advances in treatment and detection. One of the main challenges of modern medicine is detecting cancer at an early stage before it becomes difficult to treat. Despite scientific advances in the field of early cancer detection surveillance, several gaps remain. Existing diagnostic systems can be improved or redesigned. The CADUCEUS expert system was developed by Harry E. Pople Jr. in 1984. CADUCEUS is an experimental expert system for medical diagnosis [1]. The expert system for cancer diagnosis based on a reduced rule base, designed by Fatih Başçiftçi and Emre Avuçlu, was published in April 2018. It uses a simplified rule base to assess cancer risk based on the symptoms presented by an individual, thus allowing for faster results [2], the Ibex Medical Analytics expert system, Founded in 2016, this company offers cancer diagnostic solutions powered by artificial intelligence, helping pathologists ensure better care for cancer patients and The expert system for early detection of cervical cancer, Developed by Karem R. Domínguez Hernández et al., this system was published in 2013. It comprises three phases: risk assessment, interpretation of cytological images, determination of precancerous lesions, and integration fuzzy logic to improve diagnostic accuracy [3]. Thus, this work represents a significant advance in combining the advantages of existing expert systems with the current requirements of precision medicine and care adapted to the realities of different clinical contexts.

2. Literature Review

2.1. Medical Expert Systems

2.1.1. Introduction to Medical Expert Systems for Cancer Diagnosis

Expert systems are used in various fields of medicine, particularly for early cancer diagnosis, relying on rule-based approaches to assist physicians in decision-making. These systems combine clinical rules and medical expertise to perform differential diagnosis, which is crucial in detecting cancer patients, where early diagnosis can significantly improve or reduce risk.

2.1.2. Challenges and Prospects of Expert Systems for Early Cancer Diagnosis

Expert systems in the field of cancer face several challenges, including complex data management, integration with existing clinical systems, and acceptance by clinicians. However, recent advances in artificial intelligence, machine learning and expert systems open new perspectives for improving early differential diagnosis.

2.1.3. Expert Systems Based on Convolutional Neural Networks (CNNs)

Expert systems based on convolutional neural networks are deep learning models capable of automatically extracting relevant features from medical images, using convolutional neural networks, which today represent an advanced diagnostic technique in medicine. Several expert systems based on CNNs have been developed for early differential cancer diagnosis, specifically for lung, skin, and breast cancer.

2.1.4. Fuzzy Expert System

The fuzzy expert system is an innovative approach to early breast cancer diagnosis. It combines medical image analysis and clinical data with fuzzy logic to effectively manage the imprecision and uncertainty inherent in medicine.

2.1.5. Case-Based Reasoning (CBR) System

Case-based reasoning (CBR) is a domain-based approach and an artificial intelligence method that mimics how human experts in a given field make decisions based on similar past cases. In the context of breast cancer diagnosis, this approach assists physicians by retrieving relevant past cases to guide the current differential diagnosis.

2.1.6. Computer-Aided Diagnosis (CAD) Systems

In the field of computer science, there are several decision-support tools. CAD systems are computer tools designed to assist physicians in medical image analysis and clinical decision-making, particularly in the screening and early diagnosis of cancer. They do not replace physicians but provide a rapid, automated second opinion to the physician [4].

2.2. Cancer

Cancer is an umbrella term for a group of diseases characterized by the uncontrolled and abnormal growth of cells in the body. This growth can form masses called tumors, which can be benign (noncancerous) or malignant (cancerous). Cancer can affect virtually every organ and tissue in the body. Cancer cells can spread (metastasize) to other parts of the body through the blood or lymphatic system, making the cancer potentially very serious [5, 6].

The causes of cancer can be multiple, including genetic factors, environmental exposures (such as smoking, pollution, or radiation), and viral infections.

Treatment varies depending on the type of cancer, its stage, and the specific characteristics of each patient. It may include interventions such as surgery, chemotherapy, radiotherapy, or targeted treatments.

2.2.1. Type of Cancer

Breast Cancer

Breast cancer is one of the most common cancers among women. Symptoms vary depending on the size of the tumor and its location, but common signs include [6]:

Symptoms and signs:

- Palpable mass in the breast: A lump or hard area in the breast or under the armpits.
- Breast skin changes: The skin may become red, rough, or dimpled, a sign sometimes called "orange peel."
- Changes in breast shape or size: Visible changes to the breast, such as sagging, shrinking, or distortion of its shape.
- Nipple discharge: A discharge of fluid (other than milk) from the nipple, sometimes bloody.
- Breast pain or tenderness: Although pain is not always present in breast cancers, it can occur, especially when the tumor is located close to the skin.

Lung Cancer

Lung cancer is often diagnosed at an advanced stage due to unspecific or absent symptoms at first. The most common signs and symptoms include [6]:

Symptoms and signs:

- Persistent cough: A cough that does not go away or gets worse over time.
- Shortness of breath: Difficulty breathing, especially during physical activities.
- Chest pain: Pain in the chest that may be related to tumor growth or the spread of cancer.
- Bloody sputum: Coughing with blood or blood-tinged phlegm.
- Unexplained weight loss: Weight loss without any apparent cause.
- Fatigue: Weakness, even after adequate rest.

Colorectal Cancer

Colorectal cancer usually grows slowly and may not produce symptoms at first. However, when there are symptoms, they often include [6]:

Symptoms and signs:

- Changes in bowel habits: This may include persistent diarrhea or constipation and a change in the frequency of bowel movements.
- Blood in stool: Blood can make stools red or black, a sign of bleeding in the intestinal tract.
- Abdominal pain: Abdominal pain or cramping that may be constant or related to changes in bowel habits.
- Unexplained weight loss: Significant weight loss for no apparent reason.
- Excessive fatigue: A feeling of constant tiredness, which may be linked to chronic blood loss in the intestine.
- Sensation of incomplete emptying: A feeling that the intestines are not completely emptied after a bowel movement.
- This process would allow doctors to detect signs of cancer more quickly, improving the chances of cure and reducing long-term treatment costs.

3. Research methodology

The objective of this research work is to help doctors diagnose cancer through an expert system called ExpertCancer, which is capable of temporarily replacing a doctor in the event of absence. A rule-based expert system is a system that contains a set of rules used to describe certain patterns. Observed knowledge is collected and evaluated using these rules. If the rules are logically satisfied, the pattern is identified, and a problem associated with that pattern is suggested. Each particular problem (symptom and sign) may require specific treatment. These rules do not take into account the uncertainty and imprecision of human-observed knowledge and reasoning, as well as real-world knowledge, characterized by incompleteness, inaccuracy and inconsistency. The rule-based approach uses IF-THEN type rules. The IF-THEN rules take the following form: IF there is a flame, there is a fire.

3.1. Research Design

A typical rules-based system has four basic components:

1. A list of rule bases, which is a specific type of knowledge base.
2. An inference engine or semantic reasoned, which infers information or takes action based on the interaction of input (user symptoms or signs) and the rule base. The interpreter executes a production system program by performing the following match-resolve-act cycle:
   a. Matching: In this phase, the left limbs of all productions are compared to the contents of working memory. As a result, a set of conflicts is obtained, consisting of instantiations of all satisfied productions. An instantiation of a production is an ordered list of working memory items that satisfies the left side of the production.
   b. Conflict Resolution: In this phase, one of the production instantiations of the conflict set is chosen for execution. If no production is satisfied, the performer stops.
   c. Act: In this phase, the production actions selected in the conflict resolution phase are executed. These actions can modify the contents of working memory. At the end of this phase, execution returns to the first phase.
3. Temporary working memory.
4. A user interface or other connection to the outside world through which input and output signals are received and sent.

3.2. The Proposed System Architecture

Typically, patients visit hospitals to complain about their illnesses, and ExpertCancer users ask patients about their illnesses and search for symptoms and signs in the knowledge base. If the symptoms match those in the knowledge base, the user gives the prescription to the patient. The proposed framework for the ExpertCancer Expert System is shown in Figure 1. This figure presents the different modules that work together to realize a complete rule-based expert system.

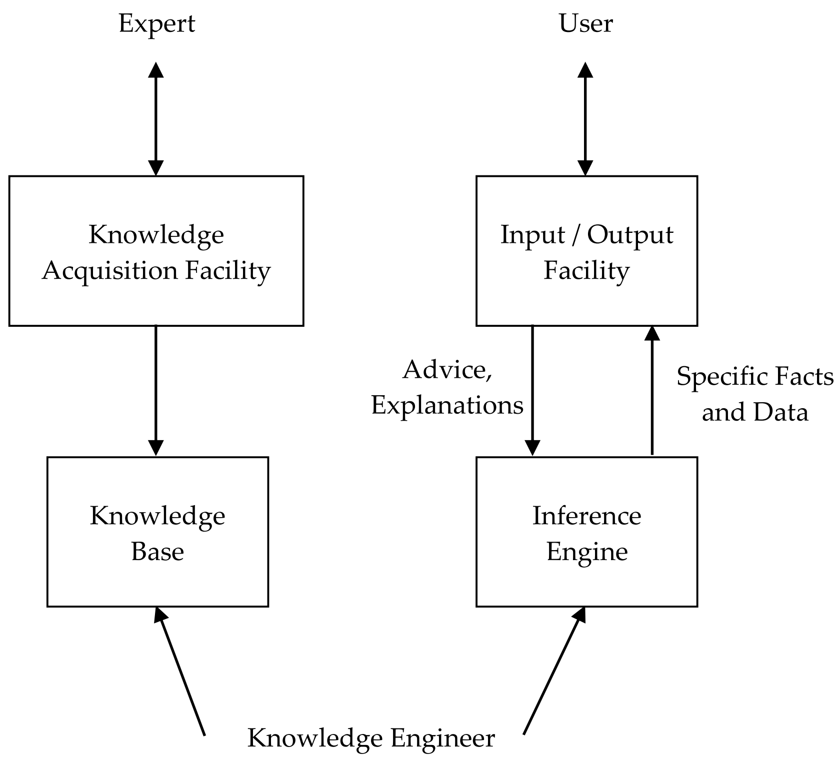

Fig. 1 Proposed architecture for ExpertCancer

This section describes the methodology phases, knowledge flows, and outputs during the application life cycle. It consists of the phases as following:

3.3. System Architecture

The expert system was developed using a three-level architecture:

- Knowledge base: This layer contains a set of rules derived from medical literature, clinical experiences and recommendations from oncologists. These rules cover information such as clinical symptoms, risk factors, results of diagnostic tests and patients' medical histories.
- Inference engine: The inference engine is responsible for applying the rules contained in the knowledge base to a patient's clinical data. When a patient enters the system with their information, the inference engine analyzes this data against the rules to generate a probability of diagnosis for the cancers concerned.
- User interface: The interface allows doctors to interact with the system. It is designed to be simple and intuitive, allowing users to enter clinical data (symptoms, tests, history) and obtain recommendations on which cancers are most likely to affect the patient.

3.4. Knowledge Acquisition Process

The knowledge base constitutes the brain of an Expert System because all the facts essential to constructing the rules are contained in the knowledge base. This knowledge constitutes the main source of rules for Expert Systems. The most important source of knowledge acquisition for the ExpertCancer Expert system was consultation with general practitioners, Internet medical websites, medical books, research articles and journals. Knowledge-based on the acquisition of disease symptoms. Knowledge is represented in the form of rules.

Table 1. Rule base for the proposed ExpertCancer expert system

Rule 1 :	If a patient presents with a palpable breast lump, then breast cancer is likely, and additional diagnostic imaging tests (e.g., mammography, ultrasound) should be performed.
Rule 2 :	If a patient presents with signs of breast skin changes (redness, dimpling, or thickening), then breast cancer is likely, and the patient should be referred to a specialist for further evaluation (e.g., biopsy).
Rule 3 :	If there is a significant change in breast shape or size without a clear explanation, then breast cancer is likely, and imaging tests should be performed.
Rule 4 :	If there is breast discharge (especially bloody or clear), then breast cancer is likely, and further investigations should be carried out with diagnostic imaging and biopsy.
Rule 5 :	If breast pain or tenderness is persistent or unrelated to menstrual cycles, then there is a possibility of breast cancer or another breast pathology, and imaging and examination should be performed.
Rule 6 :	If a cough persists for more than three weeks, then lung cancer is likely, and the patient should be referred for chest imaging (e.g., X-ray, CT scan).
Rule 7 :	If shortness of breath or difficulty breathing is present, then lung cancer is likely, and diagnostic tests should be performed (e.g., X-ray, CT scan).
Rule 8 :	If the patient reports unexplained or persistent chest pain, then lung cancer is likely the cause, and appropriate diagnostic tests should be performed (e.g., X-ray, CT scan).
Rule 9 :	If bloody sputum is produced (hemoptysis), then lung cancer is likely, and a chest X-ray should be performed, and the patient should be referred to a pulmonologist.
Rule 10 :	If the patient presents with unexplained weight loss, fatigue, or both, then consider lung cancer or another systemic condition and perform a complete workup, including imaging and laboratory tests.
Rule 11 :	If there is a change in bowel habits (e.g., diarrhea, constipation), then consider colorectal cancer and refer the patient for further investigation (e.g., colonoscopy)
Rule 12 :	If the patient reports blood in the stool, then colorectal cancer is likely, and a colonoscopy or other diagnostic tests should be performed.
Rule 13 :	If the patient complains of unexplained or persistent abdominal pain, then colorectal cancer is likely, and imaging or endoscopy should be considered
Rule 14 :	If the patient experiences unexplained weight loss or fatigue, then colorectal cancer is likely, and a complete clinical examination, including imaging and colonoscopy, should be performed.
Rule 15 :	If the patient experiences a sensation of incomplete emptying after a bowel movement, then colorectal cancer is likely, and further investigations should be performed with a colonoscopy.

3.5. Implementation with EsysCorvid

Fig. 2 Home Exsys Corvid page

Fig. 3 Home ExpertCancer page

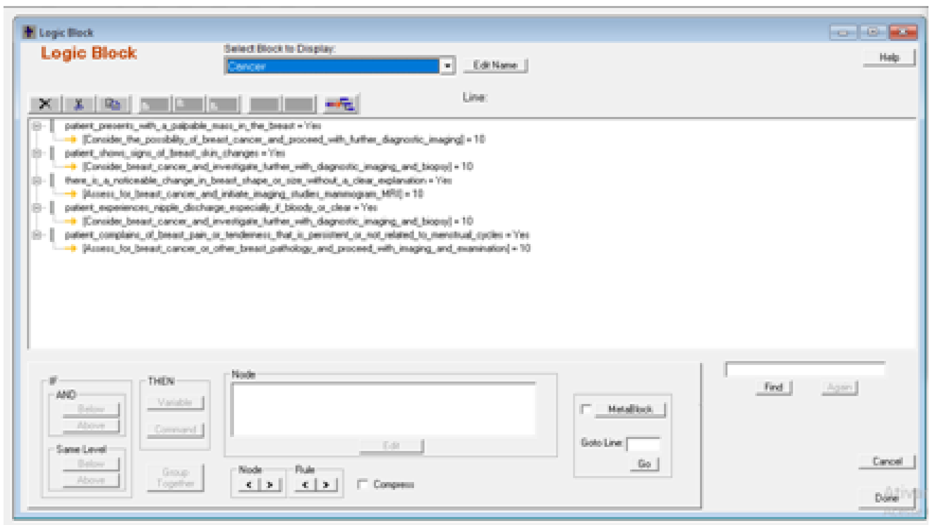

Fig. 4 Create rules with Exsys Corvid expert system

Fig. 5 Diagnosis centre for the proposed ExpertCancer

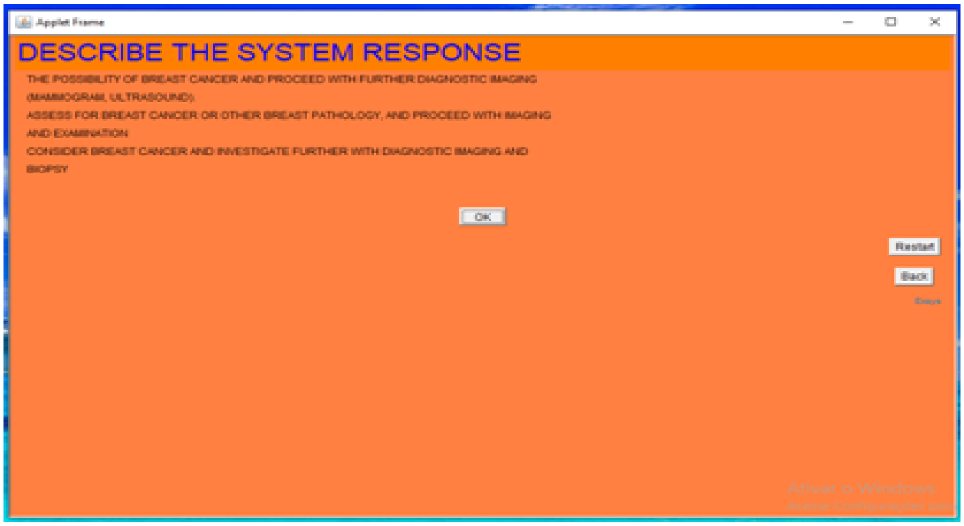

Fig. 6 Report page for the proposed ExpertCancer

4. Results and Discussion

First, the system possesses the advantages of using an expert system, which include the storage of expertise in a knowledge base, the combination of the expertise of multiple experts, its availability at all times, the ability to explain its reasoning, and the use of explicit knowledge representation. Second, the system was fully integrated with the expert system for screening and early cancer diagnosis.The development of the proposed expert system aims to assist physicians in the early differential diagnosis of cancers, particularly in settings where access to specialists is limited. Based on real medical data from anonymous medical institutions, the results obtained during the testing phases demonstrated reliable performance. The system is distinguished by the use of a rule-based inference engine combined with a regularly updated medical knowledge base. It then allows for the rapid identification of suspicious clinical signs, the formulating of a differentiated diagnostic hypothesis, and the suggestion of appropriate additional tests.

Table 2. The Kappa coefficient and contingency matrix

		Expert System	Expert
		Agree	Disagree
Expert System	Agree	7	1
Expert	disagree	2	5

Calculation of Po (proportion of observed agreement)

$$P_o = \frac{a+d}{a+b+c+d} = \frac{7+5}{15} = \frac{12}{15} = 0.8$$

Calculation of Pe (proportion of agreement expected by chance)

$$P_e = \left(\frac{(a+b)(a+c)}{n^2}\right) + \left(\frac{(c+d)(b+d)}{n^2}\right)$$

$$P_e = \left(\frac{8 \times 9}{225}\right) + \left(\frac{7 \times 6}{225}\right) = \left(\frac{72}{225}\right) + \left(\frac{42}{225}\right) = \frac{114}{225} = 0.5067$$

Where,

a+b=8 (Positive according to A)
a+c=9 (Positive according to B)
c+d=7 (Negative according to A)
b+d=6 (Negative according to B)
n=15

Calculation of Kappa :

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.8 - 0.5067}{1 - 0.5067} = \frac{0.2933}{0.4933} \approx 0.595$$

As for the performance evaluation with Cohen's Kappa Coefficient, Cohen's Kappa coefficient was calculated to evaluate the agreement between the system's decisions and the diagnoses of experienced doctors. The value of Kappa = 0.595, indicates moderate to substantial agreement between the two diagnostic methods, and according to Landis & Koch, 1977 [7], this result shows that the system can reasonably reproduce human judgments in a clinical setting compared to others such as Huang et al. (2015) obtained a Kappa of 0.78 with their fuzzy system and Abdel-Aziz et al. (2017) reported a Kappa of 0.74 in their CBR system.

5. Conclusion and Future Work

The results of this study demonstrate that a rule-based expert system is a valuable tool for early cancer detection. With a Kappa of 0.595, our system becomes more confident and provides valuable assistance to physicians, improving patient diagnostic accuracy. The results of this study highlight the importance of a rule-based expert system in cancer screening and early diagnosis. In the future, it would be relevant to integrate this system into a broader clinical setting, ensuring its regular updates and continuous adaptation to evolving medical knowledge. Integrating artificial intelligence could also strengthen this system's ability to handle more complex cases and personalize diagnosis for each patient.

References

[1] Fatih Bascftci, and Emre Avuclu, "An Expert System Design to Diagnose Cancer by Using a New Method Reduced Rule Base," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 113-120, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[2] E. Harry Pople, *CADUCEUS: An Experimental Expert System for Medical Diagnosis*, 1984. [CrossRef] [Google Scholar] [Publisher Link]

[3] Karem R. Dominguez Hernandez et al., "Development of an Expert System as a Diagnostic Support of Cervical Cancer in Atypical Glandular Cells, Based on Fuzzy Logics and Image Interpretation," *Computational and Mathematical Methods in Medicine*, vol. 2013, pp. 1-17, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[4] J. Dheeba, N. Albert Singh, and S. Tamil Selvi, "Computer-Aided Detection of Breast Cancer on Mammograms: a Swarm Intelligence Optimized Wavelet Neural Network Approach," *Journal of Biomedical Informatics*, vol. 49, pp. 45-52, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jacques Ferlay et al., "Cancer Statistics for the Year 2020: An Overview," *International Journal of Cancer*, vol. 149, no. 4, pp. 778-789, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Catherine de Martel et al., "Global Burden of Cancer Attributable to Infections in 2018: a Worldwide Incidence Analysis," *The Lancet Global Health*, vol. 8, no. 2, pp. E180-E190, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] J. Richard Landis, and Gary G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159-174, 1977. [CrossRef] [Google Scholar] [Publisher Link]