THE DEVELOPMENT OF A RULE-BASED SYSTEM FOR THE DIAGNOSIS OF CANDIDIASIS

Eteng, Idongesit Efaemiode¹, Udeze, Leonard Chinedu², Anuonye, Chinasa³ and Okon Utibeima⁴

Department of Computer Science,
University of Calabar, Calabar, Cross River State, Nigeria.
¹idongeteng@gmail.com, ideteng@unical.edu.ng, ²udezechinedu@unical.edu.ng, ³chinasaanuonye@gmail.com
⁴utibeima_okon@yahoo.com (MD)

ABSTRACT: Candidiasis is one of the most common opportunistic fungal infections in existence. Globally, about 75% of women suffer from at least one episode of vulvovaginal candidiasis (VVC) annually. The nature of genital candidiasis infections, which includes vaginal and balanitis (inflammation of the glans penis) often hinders patients from seeking appropriate medical assistance in formal health facilities because these patients usually feel shy to engage experts. We developed a Web-based Expert System that provides a self-administered diagnosis of this condition. The system provides a diagnosis in conjunction with laboratory tests verification before treatment advice and remedial measures are recommended. Requirement gathering and specification were carried out through informal interviews and interactive sessions with health care providers, visitation to the laboratory and review of relevant literature on candidiasis infections. Simple production rules, as well as coded questions and answers, were employed for inferences. Valid and reliable inferences were made regardless of the Candidiasis scenario cases used in testing the system. Fifty (50) persons were tested and an average of 12% tested positive to the five (5) various types of Candidiasis. Treatment was recommended for such patients. The system could serve as initial diagnosis support for candidiasis.

Keywords: Candidiasis, Rule-based system, Expert system, Diagnosis.

I. INTRODUCTION

Candidiasis (a fungal infection) is one of the most common opportunistic infections affecting humans. Opportunistic infections occur commonly in individuals with immune-compromised conditions such as: malnutrition, chemotherapy for cancer, Acquired Immune Deficiency Syndrome (AIDS), pregnancy, diabetes, antibiotics treatment, etc. Candidiasis is most prevalent among adolescent girls and women. Candidiasis such as thrush, or diaper dermatitis is usually predominant in the first year of life while vulvo-vaginitis is more common in adolescents and adult females. This condition can be treated and the prognosis is best if diagnosed early and prompt treatment is commenced. The field of diagnosis of candidiasis has advanced during the last decade with the advent of sequencing and proteomics technology [1]. Candida species belong to the normal microbiota of an individual’s mucosal oral cavity, gastrointestinal tract and Vagina and are responsible for various clinical manifestations from mucocutaneous overgrowth to bloodstream infection [2]. There are about 15 candida species that can cause human disease, but more than 90% of invasive candidiasis is caused by C. albicans, C. glabrata, C. tropicalis, C. parapsilosis and C. krusei, each with its own virulence potential, epidemiology and antifungal susceptibility [3]. The genus Candida albicans is responsible for approximately 50-90% of human candidiasis. Generally, the most common species are: C. albicans, C. tropicalis, C. glabrata, C. dubliniensis, C. parapsilosis, C. orthopsilosis, C. metapsilosis, C. krusei, C. famata, C. guilliermondii and C. lusitaniae [4].

The treatment of candidiasis has been faced with several
challenges which include the following: Patients especially those of the opposite sex find it difficult to divulge sensitive details to doctors and this can impair proper diagnosis. Also, the attitude of some medical personnel make patients feel judged or condemned; making them unwilling or reluctant at seeking medical help in formal health facilities.

This paper proposes a mobile-based expert system for Candidiasis diagnosis. To achieve this, firstly, a (self-support) decision support system to assist patients with initial diagnosis was developed. Secondly, rules for the occurrence of different types of candidiasis were obtained from medical doctors. These rules were written out in a form amenable for use by the rule-based system developed. Thirdly, the system was developed to be mobile-friendly in order for it to be accessible to patients no matter their level of computer literacy. This will help provide early and accurate diagnosis of Candidiasis, and prescribe appropriate medications. Providing a web-based, user friendly expert system to complement the effort of Candidiasis management especially in health institutions that lack professionals in this field, would make health services accessible to patients. The system will help to retain the skill of an expert medical doctor in case of any emergency, such as death. The system will also be useful in many hospitals, both private and government, especially in situations where the expert is not on seat. It would also be used in the laboratory as a tool for quick confirmatory tests. It can help promote awareness to users of the system on Candidiasis through the rich documentations available in the knowledge-base. Additionally, it will help provide up-to-date and reliable records that can guide health professionals in making inferences and valid conclusion. Finally, it will reduce anxiety on the part of the patients.

This work covered the diagnosis and recommended treatments for Candidiasis infections. It also has an in-built knowledgebase for documentation, references and preventable tips on the various forms of Candidiasis. This research work was limited to only five types of Candidiasis infections: Vulval Candidiasis, Oral thrush, Cutaneous (skin) Candidiasis, Candidal Balanitis and Candidal paronychia.

II. LITERATURE REVIEW

The term expert system apparently began to be replaced by the term “knowledge-based systems” in the mid-1980 to mid-1990’s [5, 6]. This name shift put less direct pressure on developers to build systems that were equivalent to experts, but also was a sign of a commercial and research shift away from expert systems and an evolution to other forms of problem solving approaches [7]. In general, knowledge based techniques can be classified into four, namely: rule-based systems, case-based reasoning, knowledge based systems and fuzzy expert systems/neural networks [8]. In this paper, we give much attention to rule-based ES, which applies production rules, and not to large training sets. Rule-based ES are usually based on two types of inference engine, namely: forward chaining and backward chaining systems.

Forward chaining is data-driven and proceeds from known data to a conclusion, while backward chaining is goal-driven [9]. Rule-based ES have been adopted for several medical diagnosis systems such as: Angina pectoris [10], myocardial infarction [11], malaria and typhoid fever [12], lassa fever [13], influenza [14], cardiological diseases [15], eye diseases [16], lungs diseases [17], memory loss [18] and viral infection [19].

Rule-based expert systems (ES) comprises of a knowledge base, inferential engine and user interface. In a rule-based ES, the knowledge base is expressed in the form of production rules. Such rules can be inferred from expertise or through machine learning methods. The inference engine is a computer program that obtains answers from the knowledge base. The rules present the knowledge in the form of premise-action pairs with an IF-THEN structure. The architecture of the rule-based system comprises of a) the Knowledge base which contains the knowledge need for understanding, formulating and solving problems, b) the Inference engine which uses the rule interpreter to provide methodology for reasoning, c) knowledge acquisition which involves gathering, transferring and transforming of techniques from experts and documents to the computer system in building the knowledge base, d) Explanation facility explains the system’s actions on how the intermediate or final solutions were arrived at and e) the User Interface for communicating with the user. The most popular tools for developing knowledge based systems are: SL5 Object and CLIPS [20].

Diagnostic ES assists in disease identification and description of treatment methods and also takes into account the user’s capability in dealing with system interaction [26]. Rule-based ES represents knowledge in the form of rules which are specifications of relations, recommendations, directives, strategies or heuristics [27]. The interface of an ES can be: Question-and-Answer, menu driven, Graphical User Interface (GUI) or natural language-based [28]. [27] and [30] designed an ES for the diagnosis of the Ebola Virus Disease. [27] reported the work as a system that was built using Prolog as the programming language for implementation. According to [29], automated systems can aid in generation of reports to guide decision-making in healthcare and managerial processes.

Among Candida spp., Candida albicansis is the most common infectious agent. This dimorphic yeast is a commensal that colonizes skin, the gastrointestinal and the reproductive tracts. The pathogenesis and prognosis of Candidiasis infections are affected by the host immune status and also differ greatly according to disease presentations [21]. There are some methods for diagnostics which include: endpoint Polymerase Chain Reaction (PCR)-based amplification, analysis of fragment length polymorphisms, isothermal amplification methods, magnetic resonance and nanoparticles, DNA hybridization, spectroscopy-based methods, antibody-based techniques etc. There are also some bioinformatics and databases used for diagnosis such as: pipelines and workflows in fungal-omics, databases and knowledge platforms devoted to fungal organisms (e.g.
UniProt, Ensembl etc.), and data integration and computationally assisted diagnostics [1].

Computer vision has proven to be instrumental in symptoms diagnosis, management and prevention of human diseases (like schistosomiasis), nutrient deficiency, pests and plant diseases. For instance, decision trees – a Machine Learning algorithm – has been used to target health interventions like promoting medical male circumcision [22]. In a future work, we intend to explore the use of Machine Learning approaches for candidiasis prediction and treatment.

2.1 Candidiasis Stages of Colonization

The major stages of colonization of candidiasis include: colonization, superficial infection, deep seated infection and disseminated infection.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Colonicisation</td>
<td>Epithelial adhesion, nutrient acquisition</td>
</tr>
<tr>
<td>2. Superficial infection</td>
<td>Epithelial penetration, degradation of host proteins</td>
</tr>
<tr>
<td>3.Deep seated infection</td>
<td>Tissue penetration, vascular invasion, immune evasion or escape</td>
</tr>
<tr>
<td>4. Disseminated infection</td>
<td>Endothelial adhesion, infection of other host tissues, activation of coagulation and blood clotting cascades</td>
</tr>
</tbody>
</table>

Fig 1: Describes the stages of Candidiasis Colonization

### III. METHODOLOGY

#### 3.1 Knowledge Engineering

**Knowledge Engineering** is the methodical construction of knowledge-bases. Since the end result is a computer program, it is natural to apply concepts from software engineering. The steps cannot be done independently, or in a simple sequence, but typically one will have to go backward and forward between them.

#### 3.1.1 Knowledge Acquisition

**Knowledge Acquisition** is the process used to define the rules and ontology required for knowledge-based systems. The phrase was first used in conjunction with expert systems to describe the initial tasks associated with developing an expert system, namely finding and interviewing domain experts and capturing their knowledge via rules, object, and frame-based ontology. Getting the right knowledge from expert domain, makes for a more efficient expert system. Knowledge acquisition was carried out through informal interviews - These interviews were done repeatedly. Write-up from journals on diagnosis, management and treatment of Candidiasis were also used. There are existing software systems that aid in identification of certain Candidiasis species and these are often upgraded to meet emerging trends. For instance, for phenotypic yeast identification, VITEK 2 system added C. auris to its recent upgrade for laboratory use [23].

#### 3.2 Rules for Inference Rule Production

The following are rules used in developing the expert system.

**a. Production Rule for Cutaneous Candidiasis:**

IF patient has patches of itchy (sometimes painful) red, moist, weepy skin AND patient has small red bumps or pustules AND patient has peeling skin (especially between the fingers) AND patient has swollen nail folds above the cuticle THEN **Cutaneous Candidiasis** is suspected!

Next include appropriate test result.

**b. Production Rule for Balanitis**

IF patient has red patch on the glans (penis tip) AND patient has swelling and blisters on the glan AND patient has accompanied by itching and burning on the glan THEN **Balanitis** is suspected!

Next include appropriate test result.

#### 3.3 Architecture of The Proposed System Framework

This system is a three-tier distributive system, which consists of the front-tier, middle-tier, and the back-end.

- **a. Front-tier:** Contains the all user interface element.
- **b. Middle-tier:** Contains data structures (inference engine algorithm)
- **c. Back-end:** Contains the database (knowledge base).

Fig. 2 illustrates the system’s architectural frame

3.3.2 Incremental Development Model

Incremental development is based on the idea of developing an initial implementation, exposing this to user
comment and evolving it through several versions until an adequate system has been developed. Specification, development, and validation activities are interleaved rather than separate, with rapid feedback across activities. Generally, the early increments of this system include the most important or most urgently required functionality. This means that the customer can evaluate the system at a relatively early stage in the development, to see if it delivers what is required. If not so, only the current increment has to be changed and, possibly, new functionality defined for later increments [24].

3.3 System Requirement Specifications
The following are specifications for the proposed system:
1. The user should have a device that is capable of running the application.
2. There should be an Internet access.
3. The System should provide a user interface (GUI) with the aim of increasing
   the usability of the system and thereby making the system user-friendly.
4. The system should be able to provide test questions for the user.
5. The system should be able to accept patients’ medical test results.
6. The system should be able to diagnose Candidiasis infection the user has, using the answers to test answer provided by the user.
7. The system should be able to provide prescription of drugs if the result shows that Candidiasis is present.
8. The output of the system should be displayed thus:
   If a patient’s result is “Candidiasis free” then, “Candidiasis free” will be the output.
If not so, “Candidiasis present” then includes prescription.

3.4 SYSTEM DESIGN
This section discusses the architectural plan of the system. To do this, this research started with the logical design, use case diagram, class diagram, activity diagram, database design, system algorithm and flowchart and interface design.

3.4.1 Logical Design
This research design covers inputs, outputs and process of the system on the various modules which makes up the system. The methodology used here, models the system as a collection of inter-related objects where each object performs a specific function and encapsulates the data and methods needed to perform its function. This research framework has four components, which are:
3.4.4 Activity Diagrams

Activity diagrams graphically represent the performance of actions or sub-activities and the transaction that are triggered by the completion of the actions or sub-actions. It is a means of describing the workflow of activities [25].

Fig 13, describes the Activity Diagram of the Proposed System.

Fig 6: Activity Diagram of the Proposed System

Fig 7: System diagnosis operation flow chart

IV. SYSTEM IMPLEMENTATION

This section discusses the details of the system implementation for the system developed in this research; the resources required in the deployment environment and a description of various modules that make up the system.

4.1 SOFTWARE DESIGN

Login page
4.2 RESULTS

To ease future maintenance, documentation is a crucial part of any developed system. Hence the documentation of this work will be presented here. This program has been sub-divided into two modules i.e., diagnose and update. The diagnose sub-module perform diagnoses on the patient based on the user data input. Diagnose has sub-modules too. The diagnose module performs diagnosis proper by calling other functions. ‘Display’ outputs the list of symptoms, while demanding that sign be provided by the medical personnel. Rules perform some if-then-else analysis that is particular to the type of Candidiasis chosen. In ‘rules’ is embedded confirm (which does some questioning) and conclusion (which provides the result using the function result). The sub-module update simply updates the knowledge-base, with relevant information or new user requirement. It calls ‘the update’ whose duty is to allow the user to write into the knowledgebase by calling the function do-the-update.

The web-based rule-based system was validated in several ways. Firstly, several tests were carried out. These include: Rigorous testing is necessary during development and maintenance to identify defects, in order to reduce failures in the operational environment and increase the quality of the operational system. The different levels of testing strategies that were applied at differing phases of system development are:

1. **Unit Testing**: Unit testing was done on individual modules as they were completed. This test was confined only to the designer’s requirements. This was done to assure that the developed system was free from bugs and reliable, different modules of the system were reviewed and tested.

Each module was tested using the following two strategies:

a) **Black Box Testing**: In this strategy some test cases were generated as input conditions that fully execute all functional requirements for the program. This testing has been used to find errors in the following categories: incorrect or missing functions, interface errors, errors in data structure or external database access, performance errors and initialization and termination errors.

b) **White Box Testing**: In this test, the test cases were generated on the logic of each module by drawing flow graphs of that module. Logical decisions were tested on all the cases. It was then used to generate the test cases in the following cases: guarantee that all independent paths have been executed, execute all logical decisions on their true and false sides, execute all loops at their boundaries and within their operational bounds and execute internal data structures to ensure their validity. For example: in validating the test for balantitis we had to check for the condition: IF patient has red patch on the glans (penis tip) AND patient has swelling and blisters on the glan AND patient has accompanied by itching and burning on the glan then we may now go on to confirm our suspicion of balantitis with tests…

2. **Integration Testing**: Integration testing ensures that software and subsystems work together as a whole. We tested the interfaces of all the modules to make sure that the modules behave properly when integrated together.

3. **Acceptance Testing**: It is a pre-delivery testing in which the entire system was tested at the client’s site on real world data to find errors.

4. **Validation Testing**: The system has been tested and implemented successfully and thus ensured that all the requirements as listed in the software requirement specification are completely fulfilled.

5. **Execution Testing**: This program was successfully loaded and executed. Few encountered errors were corrected, the system is still being maintained and will be referenced on Github shortly.

6. **Output Testing**: The output of the project was up to the expectations and the output screens present information in an easy to interpret manner.

V. **SUMMARY**

The automation of diagnosis and treatment of Candidiasis makes its management more accurate, reliable, accessible and secure. The developed system has gone through a careful re-analysis, re-design and re-implementation to become a more stable and secure system that can be used by both patients and medical personnel in the case of infection. Here the researcher assumed a dual post of a system analyst and computer programmers assistant, with focus on software engineering principles as a means of developing quality software. The
main focus of this research was the task of improving the quality and type of treatment given to patient and reducing the time of delivery. This was achieved by adhering strictly to design; in order to obtain the expected solution.

VI. CONCLUSION

Research on rule-based systems and use of machine learning techniques for diagnosis is becoming increasingly indispensable with the outbreak of pandemics like the Ebola Virus Disease and the COVID-19. According to [31], the WHO in 2020 recommended that rapid collection and testing of requisite specimens from patients meeting the requirements for the suspected case of COVID-19 is a priority for clinical management and outbreak control, and should be done under the directives of a laboratory expert. Hence, if rule-based ES with productions and datasets can be produced for the globally endemic human corona viruses like HCoV-229E, HCoV-NL63, HCoV-HKU1 and HCoV-OC43 that are prone to causing mild respiratory disease, and the zoonotic Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) with a higher case fatality rate, then Artificial Intelligence solutions would prove to be combative in ameliorating the dangers of the escalating pandemic. It is recommended that further research grants be made to such research groups that want to adopt AI, ES, health informatics and Internet Technology in the collection and transportation of specimens (like upper respiratory and lower respiratory specimen), communication with laboratories, data sharing, development of datasets, timely processing and reporting of samples, and the computerized management of diagnostic request forms.

5.2 RECOMMENDATIONS

For the automated diagnosis and treatment of Candidiasis, we made the following recommendations:

1. To improve the speed and process of diagnose and treatment of Candidiasis (with or without expert available), the system should be recommended for hospitals and other healthcare center.
2. This system stores a number of documentation on the subject that support researcher.

To maintain the functionality of the system, there should be regular orientation for new staff (health) and patient.

VII. REFERENCES


