Clifford University International Journal of Development Studies (CLUIJODS). Volume 3 (December, 2023), pp 18 - 31
www.journal.cliffordduni.edu.ng/

ISSN: 0795-0445

RULE INDUCTION ALGORITHM IN THE DIAGNOSIS AND CLASSIFICATION OF SOME COMMONLY RELATED SEXUALLY TRANSMITTED DISEASES

[1*]Odikwa, Ndubuisi Henry, [2]Thom-Manuel, Osaki Miller, [3]Uzoaru, Godson Chetachi

[1] Department of Computer Science, Abia State University, Uturu
[3]Information and Communication Center, Ignatius Ajuri University, Rivers State, Nigeria
[2]Department of Computer Science, Clifford University, Owerrinta, Abia State, Nigeria

[*]Corresponding Author. E-mail: ndubuisi.odikwa@abiastateuniversity.edu.ng

ABSTRACT
Many lives are lost today around the whole world as a result of improper diagnosis of diseases and more so confusable diseases that have commonly related symptoms. These diseases pose huge risks to human lives when it is not detected early or misdiagnosed for other diseases. Consequently, in the diagnosis of diseases using machine-learning algorithms, more of the algorithms are suitable for the diagnosis of diseases while some may not be appropriate. In this research paper, a more suitable machine-learning algorithm was proposed; which employs rules and inferences in the diagnosis and classification of diseases. The research paper employed rule induction algorithm for the diagnosis of commonly transmitted diseases of e-coli, staphylococcus, gonorrhea, syphilis and candidiasis based on 250 patient datasets collected from Federal Medical Center Owerri. The result obtained yielded a classification accuracy of 96% sensitivity of 96% and specificity of 71%.

Key Words: Rule-Induction, Machine- Learning Algorithm, Staphylococcus, Candidiasis.

1.0 INTRODUCTION:

Machine learning algorithms have been veritable tool for the diagnosis of diseases and more so in the detection of faults in machines and many other devices. Machine learning algorithm could be employed in education, medicine, agriculture, industries, and astronauts, including virtually in every facet of life.

Odikwa *et al.* (2017) defined Machine learning algorithms as mathematical model mapping methods that is used to learn or uncover underlying patterns embedded in the data. Machine learning comprises a group of computational algorithms that can perform pattern recognition, classification, and prediction on data by learning from existing data (training set).

Generally, every disease has its own unique symptoms and most disease symptomsare alike which makes it difficult to diagnose and classify them appropriately. Classification of diseases using machine learning algorithms builds a model that generally creates patterns based on the existing data of the disease symptoms. In order words, machine learning is an aspect of artificial intelligence that can learn from human beings and transform it to an expert system(] Castareda, et al., 2015, Gregory, et al., 1996, Garzotto, et al., 2005).

Some sexually transmitted diseases such as gonorrhea, syphilis, staphylococcus, E.coli, and more have related symptoms and it keep on having positive growth in a country like Nigeria

Received: July 2, 2022. Accepted: November 25, 2022.
Published by Clifford University, Owerrinta, Abia State, Nigeria

as shown in Figure1. This makes the diagnosis and classification of these sexually transmitted diseases difficult, hence the need to employ machine learning algorithm that will assist the medical personnel in the diagnosis of these ailments in the hospitals (Hong, et al., 2014).

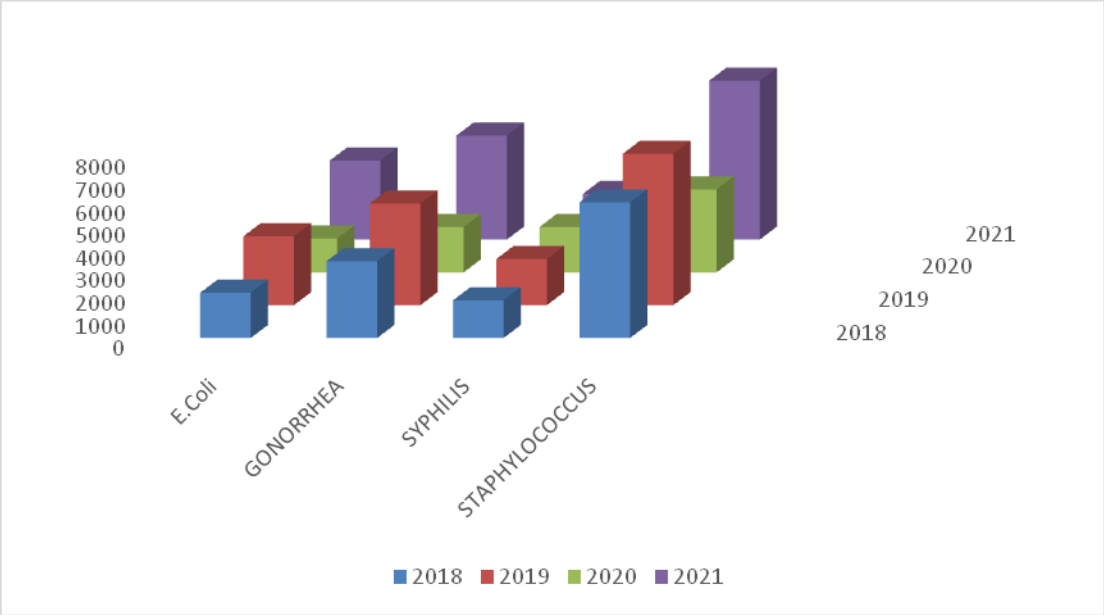

Figure.1: Common Sexually Symptom Related Transmitted Disease Growth in Nigeria[Source: Odikwa, et al., 2017)

For a Machine learning (ML) system to be useful in solving medical diagnostic tasks, such as in diagnosis and classification such asit relates to commonly sexually transmitted diseases (STD).The following features are desired: good performance, the ability to appropriately deal with missing data and with noisy data (errors in data), the transparency of diagnostic knowledge, the ability to explain decisions, and the ability of the algorithm to reduce the number of tests necessary to obtain reliable diagnosis (Gulkesen et al., 2010, Hanguang, 2012, Er, 2010).

2.0 LITERATURE REVIEW

Many researchers have made enormous contributions in the field of medicine with their research interest based on machine learning paradigm like decision trees, artificial neural network, genetic algorithm, etc to aid in medical decisions in classification and diagnosis of diseases. The fact remains that machine learning is gaining much ground in serving as expert system that will mimic human brain, as it is one of the facets of Artificial Intelligence (AI). The techniques of using machine learning is not to serve as a physician thereby replacing their work but to compliment the efforts of the physician to carry out his work effectively in diagnosing accurately with the developed model, reduce the cost of medical tests and its ambiguities and reduce his work load.

Kaustubh (2022), Manjurel (2022), defined Machine learning as the study of computation based on approaches for enhancing presentation by mechanizing the acquirement of knowledge from experience. Machine learning is a stem of artificial intelligence that involves a diversity of statistical, probabilistic and optimization techniques which allows computers to be able to learn from the past examples given to it and to be able to detect what we may call hard-to-understand patterns arising from huge, noisy data sets. It is believed that expert presentation

involves much of domain detailed knowledge, and also knowledge engineering that has resulted to hundreds and thousands of Artificial Intelligence Systems which are recently in use in most of our industries today. Machine learning is geared to providing growing levels of automation in the knowledge of engineering procedure, which in its entirety replace the much human activity that consumes a lot of time and energy or competence by detection and also by exploiting continuity in the data.

The ultimate test of machine learning is its ability to produce systems that are used regularly in industry, education, medicine, and elsewhere. Mostly the evaluation carried out in machine learning is indeed experimental in its own way, intended at presenting to show that machine learning is more reliable, efficient and purposeful than when compared with human efficiency and activity (Gazil 2021, Naresh, et al., 202).

Disease related symptoms such as commonly sexually transmitted diseases such as E.coli, gonorrhea, syphilis and staphylococcus have posed serious difficulty during diagnosis. Medical personnel spend quality time in the diagnosis and classification of these mentioned diseases and this mostly results to misdiagnosis and misclassification. Consequently, what it entails is that patients may be giving wrong medications that are not actually drugs meant for the disease they are suffering from. Misdiagnosis and misclassification of symptom related diseases have caused a lot of mayhem in our society leading to compounded health-related problems in patients, (Chi-Hua, et al.,2011, Shashikant, *et al.,* 2012, Reginald *et al., 2005*).

More so, in the diagnosis of the aforementioned diseases, doctors and other medical personnel result in sending the patients to laboratory tests which some are prone to misdiagnosis and time wasted due to sample (biomarker) culture of such diseases since they are bacteria infected diseases, Fadzil, et al. 2013, Sweta. Et al., 2018.

3.0 MATERIALS AND METHODS

A total of 250 patient records of sexually transmitted diseases of e-coli, gonorrhea, staphylococcus, syphilis and candidiasis were obtained from Federal Medical Center Owerri, located in Imo state of Nigeria. Out of the 250 data got, 50 of the data, were used in the training while 200 of the data was used in testing the system.

3.1 **Architecture of the System**

Figure 2 is the architecture of the system that trains the patient data with rule induction machine learning algorithm. It also depicts the component parts of the system and how they interact with each other. The rule induction creates the rules in form of "if then, else" statements. The inference rule is a conditional statement with two parts namely; if clause and a then clause and the inferences from the patient datasets (Biomarkers) or disease symptoms, which includes; the diseases syphilis, E-coli, gonorrhea, staphylococcus and candidiasis, are integrated in the knowledge base.

The knowledge base is a special kind of database for knowledge management, providing the means for the computerized collection, organization, and retrieval of knowledge from the database. This enables the system to detect and diagnose the disease. The support vector machine which is another machine algorithm is employed to classify the diseases of the commonly related sexually transmitted. Another component of the proposed system is the security enhancement which involves the login details.

The system does not grant access to unauthorized users, it is only those that have been given access by the administrator that will be able to login to the system after authentication. The medical doctors are the caliber of persons who are eligible to login into the system with their login parameters and thus diagnose the diseases and classify them. After the diagnosis and classification, the results are displayed and printed out.

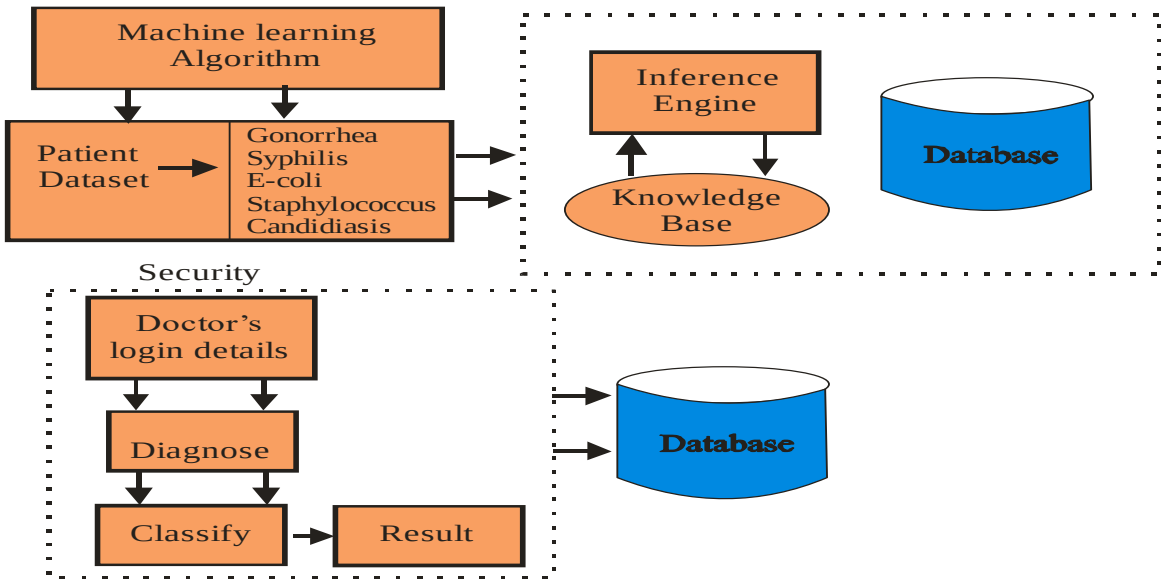

Figure. 2: Architecture of the System

3.2.**Analysis of the System**

To fully analyze the proposed system, USE CASE, Sequence diagram and collaboration diagrams which are unified markup languages (UML) as depicted in figures 3.3, 3.4 and 3.5 are employed in the analysis.

3.2.1 Detailed Algorithm of the System

The algorithm of the proposed system is as follows;

Step 1: Extract the feature vectors (symptoms)from the patient dataset.

Step 2: Create rules and inferences from the feature vectors

Step 3: Create knowledge base for the diagnosis of the disease.

Step 4: With support vector machine, classify the diseases using binary classification

Step 5: diagnose and classify the diseases

Step 6: Print patient's results.

3.2.2 The Rule-Base Formation

The rule base formation of the proposed system is based on the "if-then-else" rules using rule induction as a knowledge base in detecting diseases with discrete variable biomarkers such as syphilis, e-coli, staphylococcus and gonorrhea used for testing the model. The construction of the rule base is shown in Tables 1 and 2, 3, 4 and 5.

Table 1: Gonorrhea Rule Base

Symptoms Rules	Unusual green or yellow discharge in vagina or penis	Pain or burning when urinating	Pain in the testicles or lower abdomen	Inflammation or swelling of the foreskin	Result
1	√	√	√	√	gonorrhea
2	X	X	√	√	No gonorrhea
3	√	√	X	√	gonorrhea

Table 2: E-coli Rule Base

Symptoms Rules	Stomach cramp	Nausea	Fatigue	Diarrhea	Low fever	Result
1	√	√	√	√	√	E-coli
2	√	√	√	X	X	E-coli
3	X	X	√	√	√	No E-coli

Table 3: Staphylococcus Rule Base

Symptoms Rules	Nausea and vomiting	Sensational movement	Stomach pain	Abdominal pain	Vaginal itching	Dehydration	Painful intercourse	result
1	√	√	√	√	√	√	√	Staph
2	√	√	X	√	X	√	√	Staph
3	√	X	X	X	√	√	√	No Staph

Table 4: Syphilis Rule Base

Symptoms Rules	Painless sore	Fever	Headache	Swollen gland	Red rashes	Result
1	√	√	√	√	√	Syphilis
2	√	X	X	√	√	Syphilis
3	X	√	√	X	X	No syphilis

Table 5: Candidiasis Rule Base

Rules / Symptoms	Red rash	Vaginal itching	Pain during sexual intercourse	Abnormal vaginal discharge	Result
1	√	√	√	√	Candidiasis
2	√	X	√	√	Candidiasis
3	X	X	X	√	No candidiasis

3.3. The System Design Methodology

The system design methodology applied in this system is object-oriented analysis design methodology (OOADM). Object-oriented analysis and design (OOAD) is a popular technical approach for analyzing and designing an application, system, or business by applying object-oriented programming, as well as using visual modelling throughout the development life cycles to foster better stakeholder communication and product quality.

OOAD in modern software engineering is best conducted in an iterative and incremental way. Iteration by iteration, the outputs of Object-oriented analysis and design activities, analysis models for Object-oriented analysis and design models for Object-oriented design respectively, will be refined and evolve continuously driven by key factors like risks and business value. Object-oriented analysis and design methodology is also a waterfall model; it is a popular technical approach for analyzing, designing and application, system or business by applying the object-oriented paradigm and visual modelling throughout the development life cycles of foster better stakeholder communication and production quality, (Wikipedia 2017).

Design methodology refers to the development of a system or method for a unique situation. Today, the term is most often applied to technological fields in reference to web design, software or information systems design.

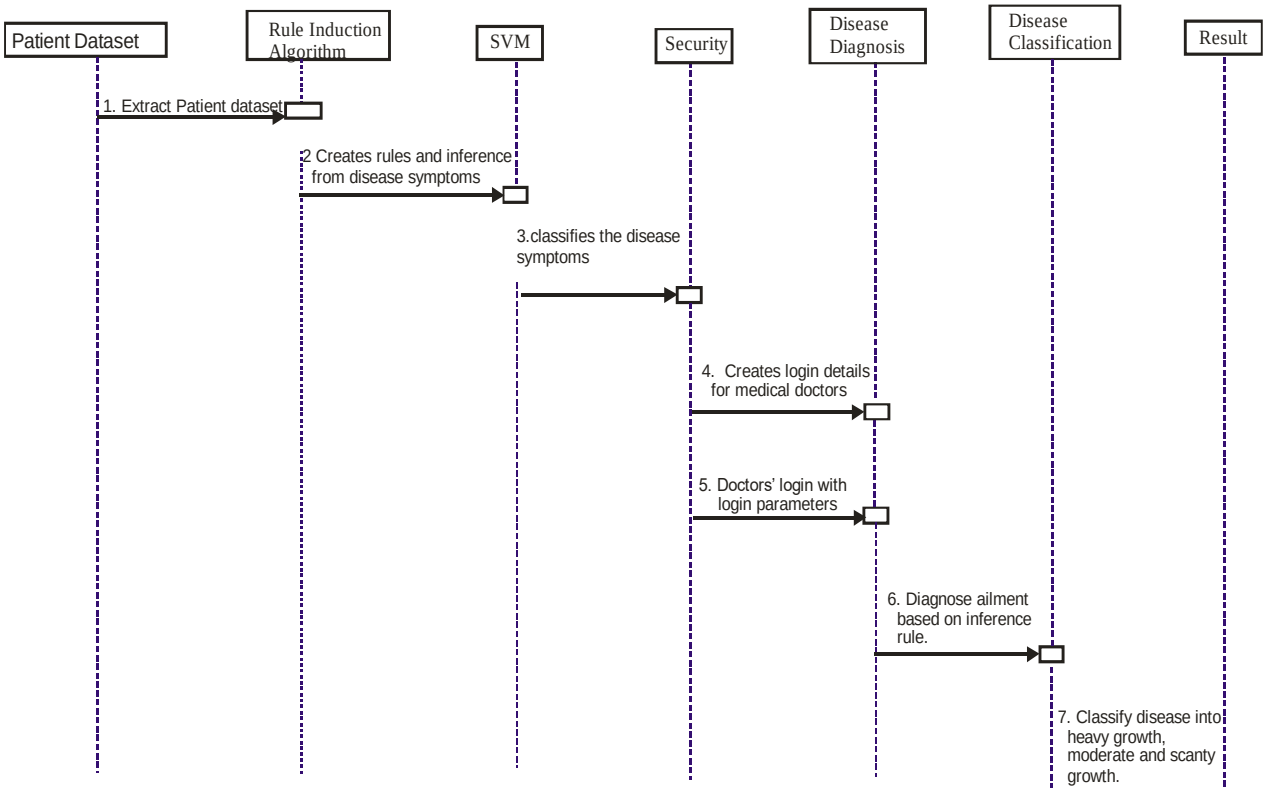

Fig. 3: The Sequence Diagram of the System

In this research, the system is analyzed using the object-oriented analysis and design methodology (OOADM). The figure 3, shows the sequence diagram of the system, a critical component from the architectural design shown in figure 2. It depicts the flow of signals from one component to another. The figures 4 and 5 are the patient-doctor and machine learning UML collaboration diagrams respectively while the figure 6 depicts the USE Case diagram displaying the various actors that interact with the system.

The USE case diagram has the actors as the patient, the machine learning algorithm and the medical doctors. These three actors create work in synergy with each other as the machine learning is employed in the training of the patient data, the patient is the one that has appointment with doctor while the doctor uses the machine learning algorithm to detect and diagnose the disease from the patient disease symptom.

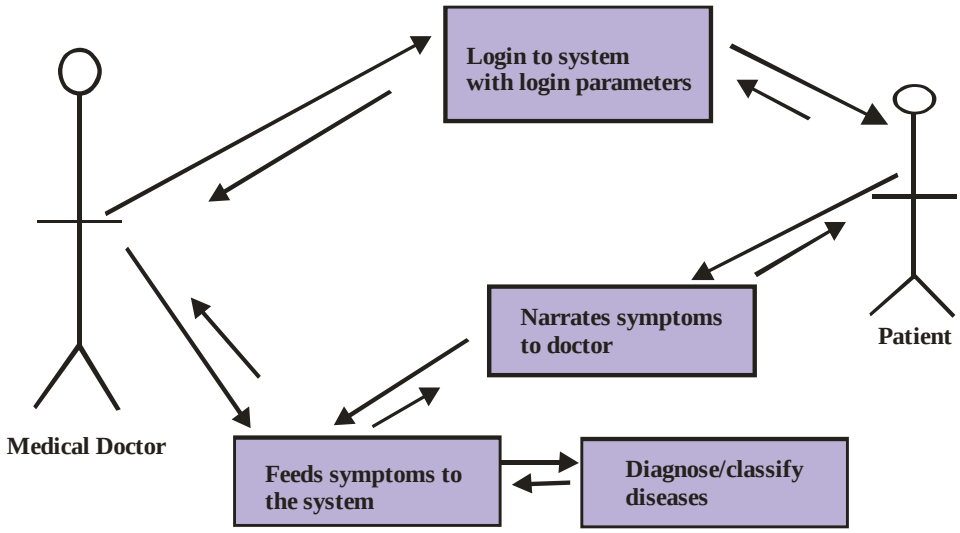

Fig. 4: Patient-Doctor UML Collaboration Diagram

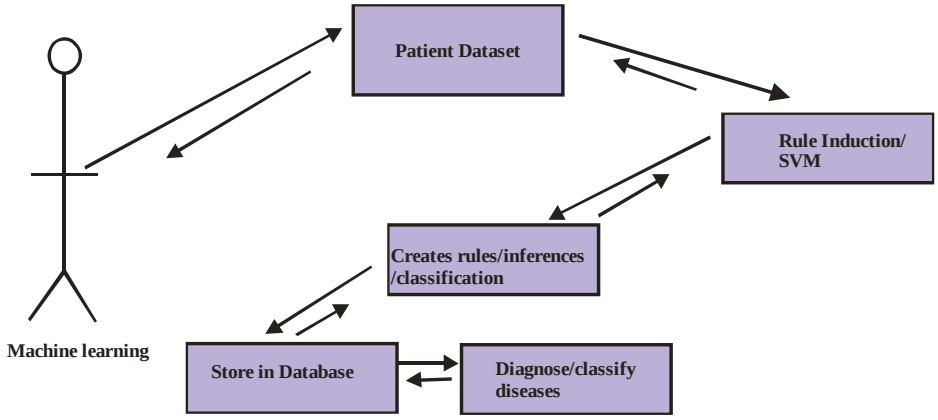

Fig. 5: Machine Learning UML Collaboration Diagram

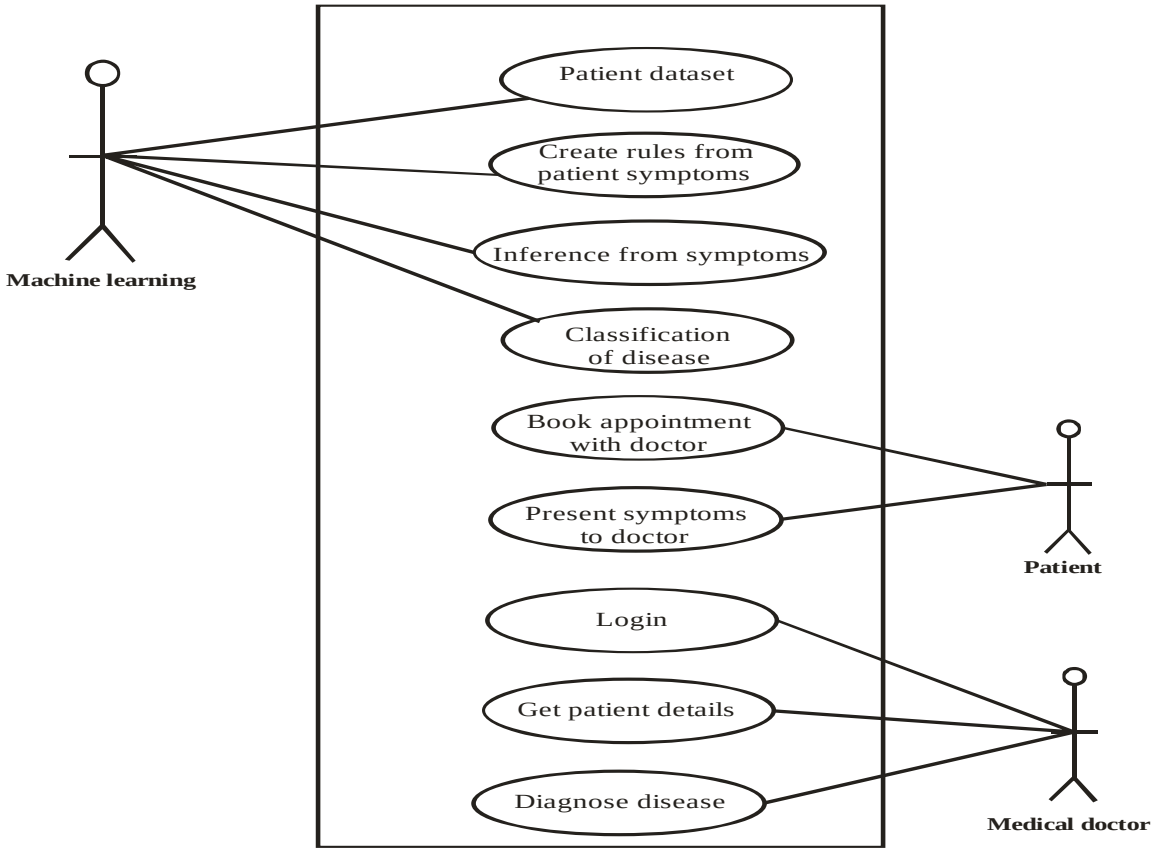

Fig. 6: The USE Case UML diagram

4.0 **Discussion of Results**

To ascertain the veracity of this system that is able to diagnose and classify commonly related sexually transmitted diseases using rule induction, many experiments were conducted in order to test the system. The user interface displayed in Figure 7 is designed based on the rule induction algorithm of if-then statements. The user interface takes cognizance of the related symptoms of commonly sexually transmitted diseases. The application is easy to use by the medical personnel as check boxes are tagged alongside the disease symptoms. This is for easy referencing in computing the result diagnosis. The Figures 8 and 9 depict the user interface showing how the system classifies diseases of sexually transmitted diseases. The classification is geared towards determining the level of the sickness in a patient. When the classification is ascertained the medical personnel apparently administers the right drug to the patient.

.

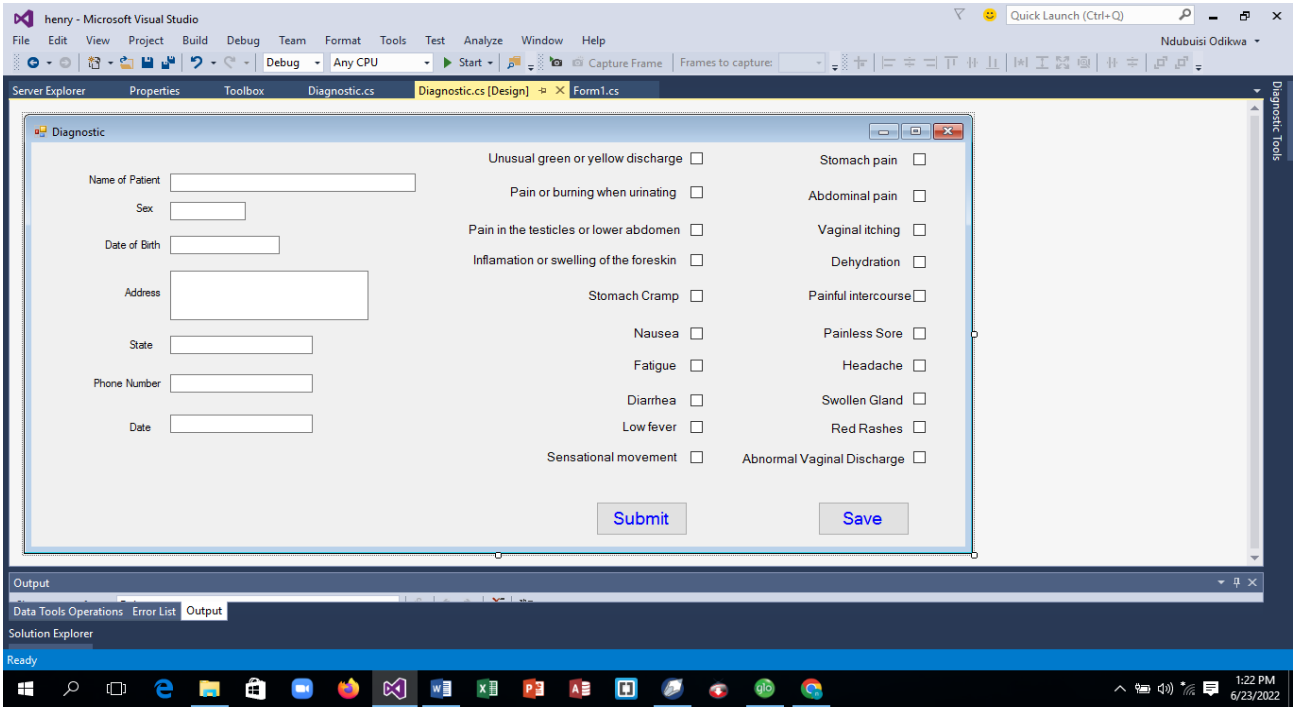

Fig 7: The Diagnosis Application

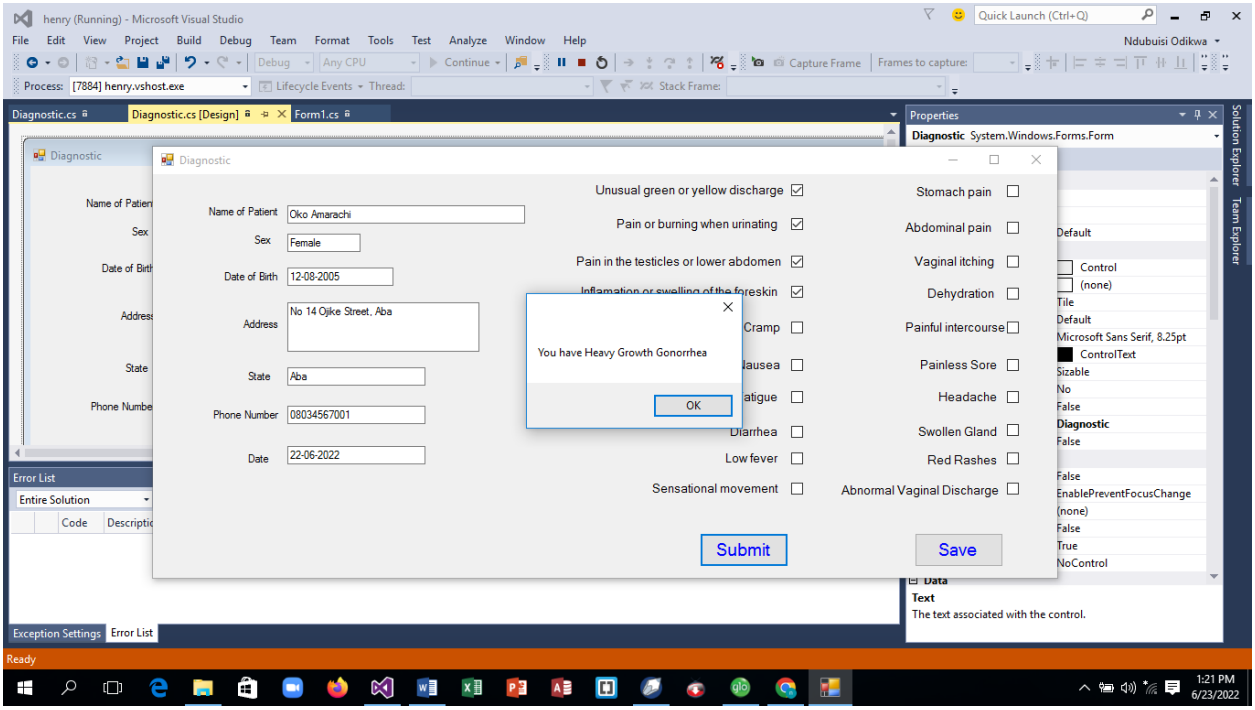

Fig. 8: User Interface Displaying Disease Classification 1

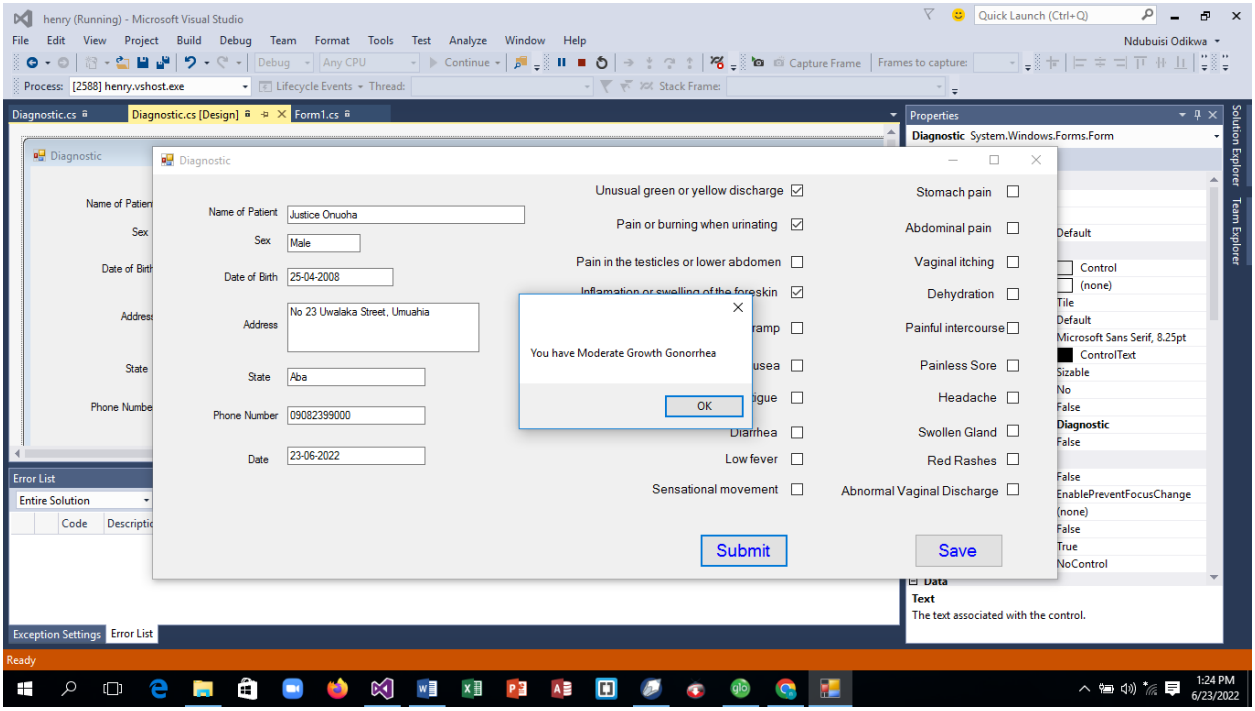

Fig. 9: User Interface showing Disease Classification 2

4. 1 Testing Accuracy of the Proposed System Using Data of commonly related diseases of gonorrhea, syphilis, candidiasis, staphylococcus and e-coli.

Testing Population = 250
TP = 235, FN = 8, FP = 2, TN = 5
Sensitivity = TP/(TP + FN)
= 235 / (235 + 8)
= 0.96
Specificity = TN / (FP + TN)
= 5 / (2 + 5)
= 0.71
Accuracy = (TP + TN) / (TP + TN + FP + FN)
= (235 + 5) / (235 + 5 + 2 + 8)
= 0.98 * 100
= 96%

Table 7: Result from Disease Classification with 200 Patient Data

Name of Disease	No of Patient	Classification
Gonorrhea	20	Moderate Growth
	15	Heavy Growth
E-Coli	17	Moderate Growth
	12	Heavy Growth
Staphylococcus	25	Moderate Growth
	15	Heavy Growth
Syphilis	23	Moderate Growth
	20	Heavy Growth
Candidiasis	25	Moderate Growth
	28	Heavy Growth

Table 8: Comparison of machine Learning Classification Accuracy in Diagnosis of STD

Machine Learning Algorithm	Performance Measure (%)
Rule Induction	96
Support Vector Machine	78
Neural Network	98
Bayes	88
Genetic Algorithm	75

In Table 7, the system designed was able to classify 20 patients' data as having moderate growth, 15 of the patients as having heavy growth of gonorrhea diseases. It also follows that 17 patients and 12 patients were classified as having moderate growth and heavy growth of E-coli respectively. Consequently, classification on staphylococcus shows that 25 patients were classifies as having moderate growth and 15 of the patients as having heavy growth while the system was able to classify 23 patients and 20 patients as having moderate growth and heavy growth of syphilis. Finally, in the classification of candidiasis diseases, 25 patients were classified as having moderate growth while 28 patients were classified as having heavy growth.

4.2 Classification Accuracy with other Classifiers

In Table 8, comparisons were made with different machine learning algorithms in classifying sexually transmitted diseases. From table 8 shown, neural network and rule induction yielded classification accuracy of 98% and 96%. Furthermore, support vector machine (SVM) has a classification accuracy of 78% while genetic algorithm has an accuracy of 75%. Also, Bayes classifier has a classification accuracy of 88%. From the result shown, it could be deduced that rule induction accuracy that was employed in diagnosing and classifying commonly sexually transmitted diseases yielded a high classification accuracy than the other classifiers previously used by researchers Therefore, rule induction is a good machine-learning algorithm for diagnosing and classifying diseases that can be detected by their symptoms.

5.0 CONCLUSION

Commonly related transmitted diseases are very difficult to diagnose. This is because they almost have the same symptoms with varying differences. The application of rule induction is a good machine learning algorithm was able to diagnose the diseases and using the diseases' symptoms classified the diseases into heavy and moderate growths. This project have many significances in the medical sector. There are confusable symptoms that exit between many related diseases such as sexually transmitted diseases. Most of the inherent significances of this paper is that diagnosis of sexually transmitted related diseases could be taken care by diagnosing with machine earning algorithms. Consequently, the classification of diseases at every stage is very important, and this research paper offers that. The significance of classification of diseases is to know how better to treat that disease by applying the proper medication and drug dosage at a particular time of the disease development. To classify diseases, machine-learning algorithms are employed especially for diseases with closely related symptoms.

REFERENCES

Castareda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, P.,…Suh, K.(2015).Clinical decision support system for improving diagnostic accuracy and Achieving precision medicine, *Journal of clinical bioinformatic* 5(4), 1-16.

Chi-Hua, C. and Semir, Z. (2011). Frontoparietal Activation Distinguishes Face and space From Artifact Concepts. *Journal of Cognitive Neuroscience, 23(*.9), 258-256.

Er, O., Termurtas, F.,&Tanrikulu, A.C.(2010). Tuberculosis disease diagnosis Using Artificial neural networks: *US National library of medicine: Journal of science*.*10*(9), 1610-1625.

Fadzil, A., Nor A.M., Zakaria H., Muhammad, K., Osman, J.(2013). Intelligent Medical Disease Diagnosis using Hybrid Genetic Algorithm; J med systems. *Springer Science and Business Media*,New York, *16*(13), 121-134.

Garzotto, M., Tomasz, M., Beer, R., Guy, H., Peter, L., Yi-Chiny, H., …Motomi, M.(2005). Improved Detectio of Prostate Cancer Using Classification and Regression Tree Analysis. *Journal of clinicalOncology*, 23(19), 432-440.

Gazil . A.E.(2021)An Evaluation of Machine-learning Methods for Predicting Pneumonia Mortality. *Journal ofArtificial Intelligence in Medicine, 9*, 130-135.

Gregory, F.C., Constantin, F.A., Richard, A., John, A., Bruce, G.B., Richard,C.,…Peter, S.(1996). Gulkesen, K.H., Koksal, I.S., Ozdem, S., and Saka, O. (2010). Prediction of Prostate Camcer Using Decision Tree Algorithm. *Journal of Medical* Science Informatics, *8*(12), 681-690.

Hananel, H., Dan, H., Larry, M., Lorraine, O., and Shimon, S. (2012). Early Diagnosis of Parkinson's Disease Via Machine Learning on Speech Data. *IEEE 27th Convention of Electrical and Electronics Engineersin Israel*. *12*(17), 10-28.

Hanguang,X.(2012). Diagnosis of Parkinson's disease using geneticalgorithmAnd support vector machine with acoustic characteristics,*Bio-Medical Engineering and Informatics(BMEI), 2012 5th International Conference*, *12*(20), 12-29.

Hong, C., Zhibin, P., Luoqing, L., &Yoanyan, T. (2014). Error Analysis of Coefficient Based Regularized Algorithm for Density Level Detection. *Journal of MIT, 25*(4), 1107-1121.

Kaustubh, A.B., Jimmy, S., Yaser, D.A, Oh-Yong, S … Ali, K.B. (2021). Medical Diagnosis using Machine Learning: A Statistical Review. Computers, Material and Continus. 67(1), 107-125.

Manjurel, M.A., Shahana, A.L and Zahed, S. (2022). Machine Learning-Based

Disease Diagnosis: A Comprehensive Review. Health Care MDPI.

Naresh, K, Nripendra, N., Deepali, D., Kamali, G., and Jatin, B. (2021). Efficient Automated Disease Diagnosis using Machine Learning. Journal of Health Care Engineering. 1(12), 1-13.

Odikwa, H.N., Ugwu, C., and Inyiama, H. (2017). Improved Decision Support System. International Journal of Artificial Intelligence and Application. 6(8), 37-55.

Reginald, A and Patoli, A.Q. (2005) Syndromic Management of Sexually Transmitted DiseasesUsing Dynamic Machine Learning and Path-Finding Algorithms2005 International Conference on Information and Communication Technologies. IEEE

Shashikant, U.G., and Ashok, A.G. (2012). Heart Diseases using Machine Learning Algorithm. Proceedings of the international Conference on Information Systems Design and Intelligent Applications (INDIA 2012). 217-225.

Sweta, G and Rajashekharaiah, K.M.M. (2018). Chronic Diseases Diagnosis using Machine Learning. International Conference on Circuits and systems in Digital Enterprise Technology.