SOLENT UNIVERSITY

FACULTY OF BUSINESS, LAW, AND DIGITAL TECHNOLOGIES

MSc Applied AI And Data Science

ACADEMIC YEAR 2021-2022

KINGSLEY CHIJIOKE OGAM-OKAFOR

STUDENT NO. Q15822630

DEVELOPING NOVEL MACHINE LEARNING APPROACHES TO PROCESS MEDICAL INFORMATION

Supervisor: Dr Shakel Ahmad

September 2022

This Dissertation is submitted in partial fulfillment of the requirements of Solent University for the degree of MSc Artificial Intelligence and Data Science

ACKNOWLEDGMENT

My innermost appreciation goes to God Almighty for giving me life, good health, protection, financial and intellectual capacity to accomplish this academic challenge.

This work was completed through the support of two major people. On the first part, I hereby acknowledge without reservation, my unwavering gratitude to my supervisor, Dr. Shakel Ahmad, of faculty of Business, Law, Digital Technology of Solent University, Southampton, United Kingdom who made this work possible. Apart from making out time despite his very tight schedule, to direct, teach and supervise me, he during my studentship motivated me not to give up on this program.

On the second part, I hereby express my deep-felt gratitude to my beloved brother, Nicholas Chibuikem Ogam for his uncommon love to me and his unwholesome financial support to me. Without his firm encouragement, I could not have enrolled for this master's degree program.

I also wish to extend my gratitude to my colleagues in the class of September 2021 set for their critical intellectual reactions in class which helped me to learn a lot and as well improved on this work.

I will not forget to mention the affection and moral support of my better half, Rosemary. Her assistance to me pushed me to complete this work at the recorded time.

ABSTRACT

This research project aims to develop novel machine learning approaches to process medical information. The amount of data created about patients in the healthcare system is always increasing. The human review of this enormous volume of data derived from numerous sources is expensive and takes a lot of time. Additionally, during a patient visit, doctors write down the patient's medical encounter and send it to nurses and other medical departments for processing. Often, the doctor doesn't have enough time to record every observation made while examining the patient and asking about their medical history which takes time for a medical diagnosis to be made. The manual review of this vast amount of data generated from multiple sources is costly and very time-consuming. It brings huge challenges while attempting to review this data meaningfully. Therefore, the goal of this research is to create a system that will address the issues. The suggested method extracts information from the voice data of medical encounters and converts it to text using Deep Learning (DL) and Natural Language Processing (NLP) techniques. More so, the system developed will improve medical intelligence processing by using deep learning to analyze medical datasets and produce results of a diagnosis, assisting medical professionals at various levels in making realistic, intelligent decisions in real-time regarding crucial health issues. The system was designed using the Object-Oriented Analysis and Design Methodology (OOADM), and the user interfaces were put into place utilizing Natural Language Processing techniques, particularly speech recognition and natural language comprehension. Speech recognition enables taking text notes, which can drastically cut down on the amount of time medical staff spends on labor-in the tensive clinical recording. By extracting different pieces of data for medical diagnosis and producing results in a matter of seconds, a deep learning algorithm demonstrates a significant capacity to construct clinical decision support systems. Deep learning solutions allow healthcare organizations to deliver personalized patients' medical history, symptoms, and tests. Natural language processing provides insights from free-text medical information for most relevant medical treatments. Thus, the type of deep learning used in this dissertation is NLP. The system's results demonstrate that the deep learning algorithm enabled medical intelligence to be 96.7 percent accurate.

ACRONYMS

AI	Artificial Intelligence
API	Application Programming Interface
CBR	Case-Based Reasoning
CRT	Cardiac Resynchronization Treatment
CNNs	Convolutional Neural Networks
DW	Data Warehouses
DL	Deep Learning
DBN	Deep Belief Networks
EHRs	Electronic Health Records
GAN	Generative Adversarial Networks
HER	Health Electronic Record
LSTMs	Long Short-Term Memory Networks
MLPs	Multilayer Perceptrons
ML	Machine Learning
NLP	Natural Language Processing
OOADM	Object-Oriented Analysis and Design Methodology
OOD	Object-Oriented Design
OTPs	One-Time Passwords
PHP	Hypertext Preprocessor
RBFNs	Radial Basis Function Networks
RNNs	Recurrent Neural Networks
RBMs	Restricted Boltzmann Machines
SOMs	Self-Organizing Maps
SQL	Structured Query Language, Or SDL
UI	User Interface

1. INTRODUCTION AND BACKGROUND

1.1 **Introduction and Background**

To generate and deliver services to patients, all actions and activities carried out in the healthcare industry are combined into medical information processes. Given the significance of medical information processing, attempts to enhance procedures can significantly affect the performance of the industry, boosting efficiency, decreasing time wasted, and opening prospects for income generation, especially in developing nations like Nigeria. When we look at the healthcare industry, we see that the availability of electronic health records in healthcare facilities produces a ton of data that is very beneficial for performing clinical research. Health electronic record (HER) systems have become widely employed in hospitals and clinics during the past few years. The analysis of this enormous amount of data will lay the groundwork for better patient care. However, it is expensive and takes a lot of time to manually review this enormous amount of data that was created from many sources. When attempting to review this data meaningfully, it poses enormous obstacles. In underdeveloped nations like Nigeria, doctors document patient visits on paper and pass them along to nurses and other medical departments for processing. These documents, which are used to process medical data for the healthcare industry, must be error-free. In hospitals, patient record errors are a major issue. During a patient visit, 81 percent of doctors did not have access to all the information they needed, according to observational research done in a university clinic (Tang, 2014).

As a result, artificial intelligence (AI) tools are playing an increasingly significant role in improving clinical research and care. Natural Language Processing (NLP) and Deep Learning (DL) approaches have been used to extract information from various HER that are locked in clinical narratives. While reinforcement learning techniques can be utilized in the context of robotics-assisted operations, computer vision techniques can be used for medical imaging, and NLP can be used to analyze unstructured data in the HER. While evaluating text and figuring out the grammatical relationships between phrases, NLP algorithms can be utilized to find clinically relevant phenotypes. The identification of a significant portion of real cases and a high positive predictive value in clinical records can be achieved using rule-based NLP approaches.

The goal of natural language processing (NLP) is to infer meaning from words by analyzing speech and text. Deep learning methods called recurrent neural networks (RNNs) are essential for processing sequential inputs including language, voice, and time-series data (Esteva, 2019). Deep learning is a kind of machine learning that can create unsupervised models from unlabeled or unstructured data. On the other hand, deep learning will be able to automatically extract the best features from the data that is provided. The practice of utilizing computer algorithms to recognize essential components of a common language and extract meaning from unstructured spoken or written data is known as natural language processing. Artificial intelligence, computational linguistics, and other machine learning fields are needed for NLP (Bresnick, 2020). To categorize patients into a subgroup according to rules and learners, NLP algorithms first extract information or concepts from HERs, then analyze extracted information. Basically, we may group the users of the healthcare information originating from the following four sources:

1. Doctors

2. Patients

3. Medical assistants

4. Pharmaceuticals

The diagnosis of an illness determines the course of every post-process. A person can receive the appropriate treatment if the ailment is correctly identified. Due to delays in decision-making, patient situations can occasionally become dangerous. A large-scale network that will accept a range of input data kinds, such as text, image, audio, time-series data, etc., is going to employ deep learning to acquire useful features for each data type in its lower-level towers. The Deep Neural Network is then able to draw conclusions based on reasoning and evidence from various sorts of data as the data from each pillar is combined and passes through higher layers. Information extraction, unstructured data to structured data conversion, document categorization, and summarizing are all areas where natural language processing can benefit the healthcare industry. Through correct prior authorization approval and effective billing, it will ultimately lower administrative costs. Additionally, it will add medical value by supporting poor clinical judgment, streamlining the evaluation of medical policy, etc (Rangasamy, 2018). In addition, it will enhance patient contact with healthcare professionals and health electronic records, raise patient health awareness, enhance the standard of care, and identify individuals who require urgent medical attention. Therefore, the goal of this research is to create a system that doctors may use to document the patient's medical inquiries while they are there. As the doctor speaks the prescriptions, natural language processing converts them to text and stores them in the database. Additionally, this will be web-based and coupled with an electronic health record, making it mobile-friendly.

1.2 Justification

The amount of patient-focused data created in the healthcare system is continually increasing. Numerous departments and/or units are being added to hospitals. Medical equipment, lab findings, electronic prescriptions, therapeutic decisions, and clinically observed values by doctors and nurses could all produce patient-oriented data. These data are dispersed, so getting access to them requires requesting one of the many hospital departments. The human review of this enormous volume of data derived from numerous sources is expensive and takes a lot of time. When attempting to review this data meaningfully, it poses enormous obstacles. Additionally, during a patient visit, doctors write down the patient's medical encounter and send it to nurses and other medical departments for processing. These documents, which are used to process medical data for the healthcare industry, must be error-free. Often, the doctor doesn't have enough time to record every observation made while examining the patient and asking about their medical history. In hospitals, patient record errors are a major issue. Any omission-related recording inaccuracy in a patient's medical investigation may have an impact on how well a patient responds to therapy. To create a medical information system, deep learning and natural language processing must be used.

1.3 Research Questions

These inquiries are addressed by this study:

1. What are the difficulties with physician clerkships and how do they affect the management of patient healthcare?
2. What are the difficulties in storing and retrieving patient data, and how does it affect the patient's medical information system?
3. How can the management of medical information systems be aided by deep learning and natural language processing?
4. What are the advantages of Natural Language Processing (NLP) and how might it help doctors achieve automatic voice-to-text conversion during patient consultations?

1.4 Research Aims and Objectives

This dissertation aims to develop novel machine learning approaches to process medical information.

The aims can be broken down in the following objectives:

1. To assess the current medical information systems and identify their limitations.
2. To create a system that can transform audio data from doctor visits into text using Natural Language Processing (NLP) and Deep Learning (DL) methods.
3. To assess the suggested system's performance in terms of data correctness and processing speed.

1.5 Significance of the Research

The analysis of deep learning and natural language processing systems in the context of medical information processing will be extremely beneficial to doctors, the government, hospital administrators, and patients.

i. The medical professionals will have access to an intelligence system that will support their clinical work and enable them to diagnose and cure patients' illnesses more swiftly and efficiently. Also, because Natural Language Processing (NLP) will translate the speech to text, the doctor won't have to write down the observations made during medical encounters with patients.

ii. By processing data in real-time, the hospital administrator will be able to eliminate duplication and retrieval delays.

iii. As patients get speedier and more precise access to healthcare, their ailments will be better.

iv. The government will profit from the new system because of the healthcare system's significant improvement, which will result in a prosperous and healthy country.

1.6 Work Program

This range includes the following three locations in detail:

i. Creating new digital medical service patterns and service process standards using the full medical information system of today.

ii. The combination of service sharing across medical institutions in a particular area together with a monitoring and evaluation system for the sharing services, as well as the integration of a national healthcare sharing platform to produce a uniform procedure.

iii. The provision of services, including patient follow-up visits, remote access to medical records and test results, and online medical consultations.

1.7 Organization of the Report

Five chapters make up the format of the research report. An overview of the project is given in Chapter one, along with a description of the problem, the goals of the research, and the importance and parameters of the investigation. A relevant literature review on the medical information system, deep learning, and natural language processing is presented in Chapter two. The study approach and the simulation tools are discussed in chapter three. Design and execution are covered in chapters four and five, simulation results are covered in chapters six along with a conclusion and a review of the research findings.

2. LITERATURE REVIEW

This chapter conducted a review of the theoretical and associated literature on deep learning, processing of medical information, and natural language processing. Reviewing some linked literature to medical records, highlighting its accomplishments and knowledge gaps, was done. The accomplishments of earlier studies and the identified research gap are demonstrated in a summary of related literature that was provided.

2.1 Processing of Medical Information

The primary goal of hospitals is to ensure the health and well-being of their patients. Professionally qualified physicians and nurses that possess not only academic expertise but also compassion for their patients are essential to providing quality medical treatment. Additionally, it calls for top-notch tools and facilities. Since it serves as a guide for the creation of the healthcare plan, good record keeping describes the most fundamental aspect of medical care. It requires accuracy and accurate information, exactly as the patient provides it. A thoroughly updated report based on this data advises medical professionals on the course of treatment to be taken (Sheridan, 2013). Inaction leads to a patient's facts and symptoms being miscommunicated and incorrectly diagnosed, which could have catastrophic implications, including the patient's death. Before keeping a patient's record, a health care professional must always be extremely cautious because many tiny errors could go wrong. Health reports, research, and statistical reports can all be derived from a patient's health record, which indicates the hospital's responsibilities in their life. If the patient is to be protected, related records such as X-rays, samples, drug records, and patient registers must also be well-maintained (Sheridan, 2013). A good record-keeping practice guarantees that the storage area is clear of extraneous and unclear information and that only the data pertinent to the patient's health is stored. The staff can easily access the information, saving time and resources.

2.1.1 Types of Records that must be kept in Hospitals

The outpatient department is typically the first stop for patients who arrive at the hospital. Here, patients initially encounter the hospital staff, who hand them a basic hospital document known as the hospital card. The patient's name, age, sex, marital status, and place of residence (but not their whole address) are listed on this form (Boonstra, 2017). On the patient's initial visit to the hospital, this file will still be available to them. No matter where they live, patients can go to any hospital or medical facility in the nation. The nurse could also record case notes in writing. Following the conversation, the nurse evaluates the situation and decides whether to treat the patient herself or refer them to the treating doctor. After that, the doctor will treat the patient and document it in the patient's file. The patient's hospital card will contain handwritten prescriptions for medication. The patient uses this card to purchase the medication at the pharmacy after being released from the hospital. If the patient is admitted to the in-patient ward, the medication is purchased from the hospital pharmacy and is charged to the patient's account at the end of treatment.

1. **Case notes for patients**: Most healthcare professionals utilize case notes, which are brief pieces of paper, to record pertinent details before entering them into the patient's medical record (Boonstra, 2017). Basic details including the signs and symptoms, findings from diagnostic tests, temperature, and blood pressure could be included. They might also contain intricate details including patient histories and allergies, as well as documentation of surgeries and other types of medical care.
2. **Patients Indexes and Registers:** By using indexing, computer-based hospital filing systems make it simple to establish patient registers. Indexing might be done quite simply by categorizing patients under various addresses and locales, as is the case in most modern countries (Kassell, 2018).
3. **Pharmacy and Drug Records:** The prescription and supply of medications creates a range of records, including pharmacy stocks, ordering, and dispensing records, requests for medications from wards and departments, records of drug administration, and prescriptions for specific patients (Kassell, 2018).
4. **Nursing and Ward Records:** The head nurse is responsible for providing letters, reports, meeting minutes, and all other relevant documents in an executive role for every institution (Sheridan, 2013). Figure 1 below shows the management organization of a hospital records service:

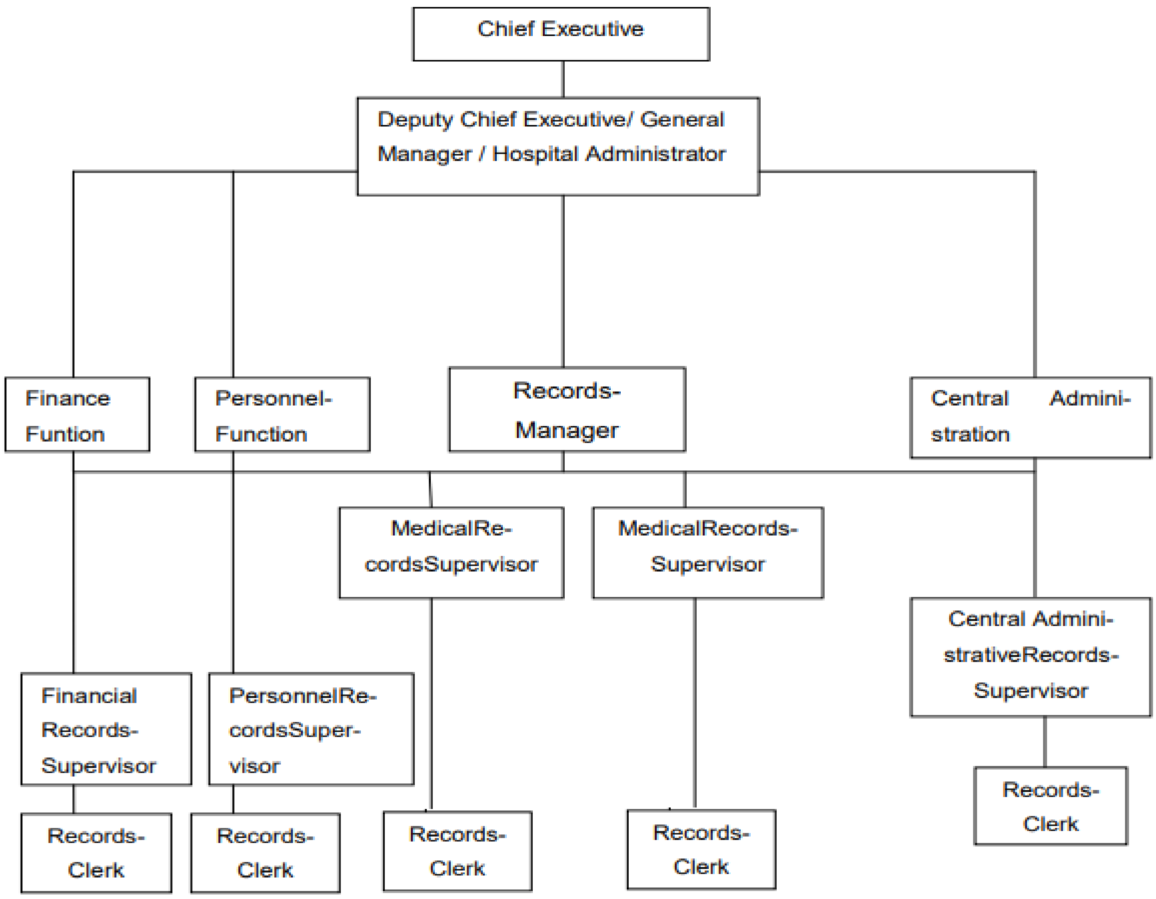


Figure 1: shows how a hospital record service is managed (Sheridan, 2013)

2.1.2 Conceptual Review of Medical Records

Any document that provides comprehensive information regarding a patient's history, clinical findings, diagnostic test results, pre-and post-operative treatment, patient progress, and medication is referred to as a medical record (Bali, 2011). It can also be characterized as a written, sequential narrative of a patient's examination and treatment that contains the patient's medical history and complaints, the doctor's physical findings, the outcomes of diagnostic tests and procedures, as well as drugs and therapeutic methods. It consists of a written chronologic narrative of the patient's first complaint(s), medical history, physical findings, outcomes of diagnostic tests and procedures, any therapeutic medications or treatments, and subsequent developments throughout the illness (Farlex Partner Medical Dictionary, 2012). Documents produced because of patient care are included in medical records.

According to Rose (2019), a patient's identifying information, such as name, date of birth, Social Security number, address, contact information, insurance information, emergency contact information, etc., must be included in a medical record at a minimum. Beyond these fundamentals, she continued, the medical record must also have enough clinical documentation that supports the need for the procedure, such as:

1. A narrative description of the patient's current condition, such as the main complaint or the basis for seeking diagnosis or therapy.
2. The objective, repeatable, and traceable facts concerning the patient's condition are included, as are the vital signs, lab results, and other physical exam findings.
3. The note's medical diagnosis and the time and date of its writing.
4. The treatment schedule, subsequent actions, and follow-up.

2.1.3 Medical Case-Based Reasoning Systems

In many fields, case-based reasoning (CBR) has proven to be an effective strategy for knowledge-based systems; however, using this technique in the medical field presents additional challenges. Case-based reasoning refers to the process of understanding and resolving new problems using examples from the past. A case-based reasoning approach tries to adapt its solutions to the present case by recalling previous cases that were like the current issue. The fundamental presumption is that similar issues will have comparable solutions. This presumption holds for many practical domains, albeit it is not always true. CBR involves two primary tasks: The first step is retrieval, which involves looking for or tallying up the most comparable cases. A sequential calculation is feasible if the case base is relatively small; otherwise, quicker non-sequential indexing or classification techniques (such as nearest neighbor match) should be used. Since a lot of research has been done in this area recently, it is now simple to identify powerful case-based reasoning (CBR) retrieval algorithms that are suitable for almost every application situation.

The second step, adaptation (reuse and revision), is adapting solutions from earlier, analogous circumstances to meet a more recent situation. A straightforward solution transfer is sufficient if there are no significant variations between the current situation and a comparable scenario. Sometimes only a few changes are needed, while other times the transition is a very difficult process. The adaptation is still entirely domain-dependent because no universal adaptation algorithms or procedures have yet been created.

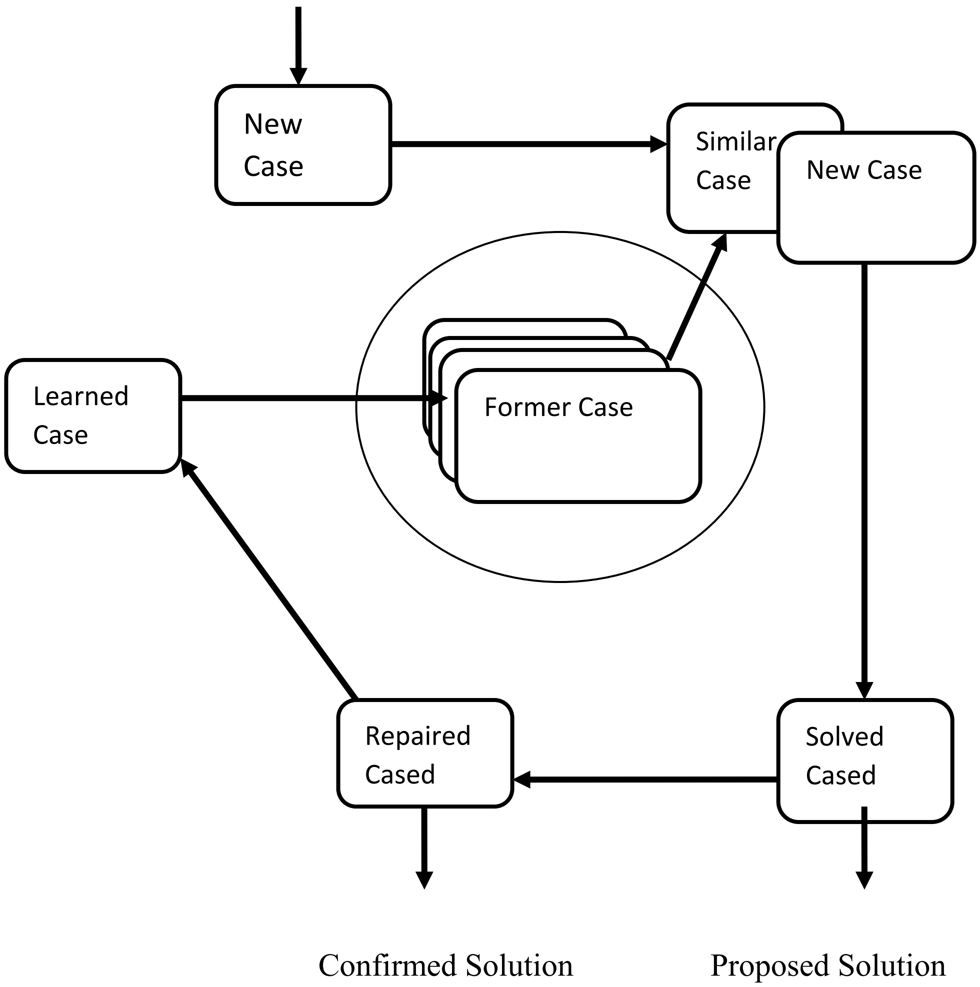


Figure 2: Cycle of Case-Based Reasoning (Aamodt, 2014)

Case-based reasoning (CBR) has primarily been used in medicine for diagnostic and in certain cases for therapeutic functions. Case-oriented strategies for tutoring and retrieval methods to look for similar photos are similar techniques that have been employed in other fields.

CASEY is one of the first medical expert systems to employ CBR methods (Koton, 1988). It addresses the diagnosis of heart failure.

2.2 Processing of Natural Language

A branch of computer science called "natural language processing" (NLP) focuses on applying computational methods to learn, comprehend, and create material in human language (Hirschberg, 2015).

Information extraction, which converts text's unstructured data into structured data (Jurafsky, 2018); conversational agents, which facilitate human-machine communication (Hirschberg, 2015); or machine translation, which uses computers to speed up language conversion to facilitate human-human communication, are some examples of NLP applications (Hirschberg, 2015). According to Hirschberg (2015), the following are the factors that have made it possible for NLP to evolve during the past 20 years:

1. A boost in computer power

2. The accessibility of vast linguistic data,

3. The creation of effective machine learning techniques, and

4. A deeper comprehension of the structure of human language and how it is used in social contexts.

2.2.1 Natural Language Processing in Intelligent Healthcare

A healthcare system is known as "smart healthcare" makes use of cutting-edge technologies like artificial intelligence (AI), blockchains, big data, cloud/edge computing, and the internet of things (IOTIoTo create a variety of intelligent systems that connect healthcare participants and improve healthcare quality (Tian, 2019). The public, healthcare service providers, and third-party healthcare participants are the three main groups of participants in smart healthcare. Representative smart healthcare scenarios peraboute participants include smart homes, smart hospitals, intelligent life science research and development, health management, public health, rehabilitative therapy, etc. The main players, cutting-edge technology, and illustrative scenarios of smart healthcare are shown in Figure 3.

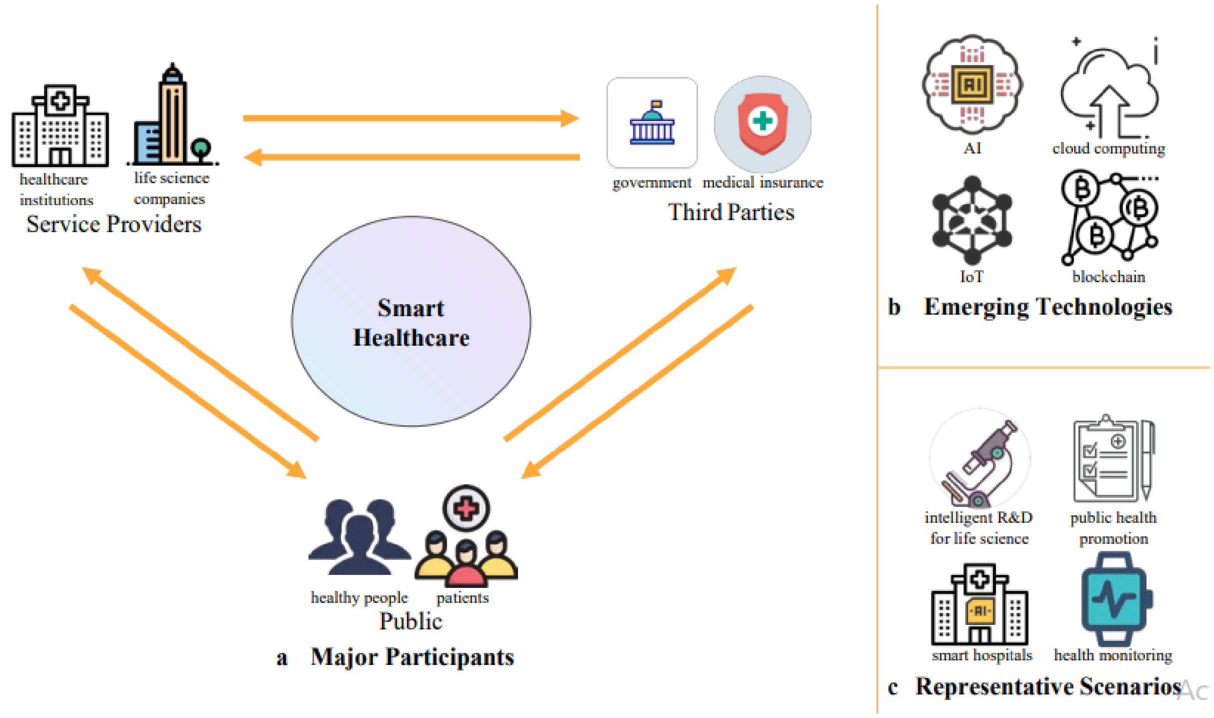


Figure 3: Intelligent healthcare (Tian, 2019)

The public, healthcare service providers, and third-party healthcare participants are the main stakeholders in smart healthcare, as shown in Figure 3a. Figure 3b shows how cutting-edge technology like artificial intelligence, blockchains, cloud computing, the internet of things, and others enable smart healthcare applications. Figure 3c shows an example of a smart healthcare scenario, which also includes intelligent research and development for life science, the promotion of public health, smart hospitals, health monitoring, and other things. Computer science and artificial intelligence's field of natural language processing (NLP) is concerned with the automatic analysis, representation, and comprehension of human language (Young, 2018).

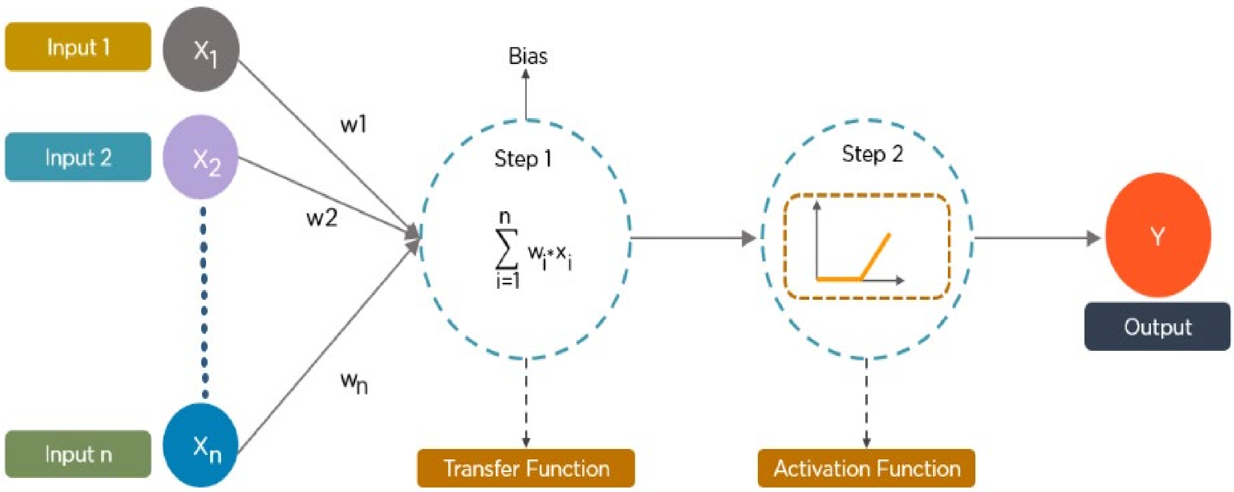
2.3 Deep Learning

Artificial neural networks are used in deep learning to carry out complex calculations on vast volumes of data. It is a form of artificial intelligence that is based on how the human brain is organized and functions. Machines are trained using deep learning algorithms by learning from examples. While self-learning representations are a hallmark of deep learning algorithms, they also rely on ANNs that simulate how the brain processes information. To extract features, classify objects, and identify relevant data patterns, algorithms exploit unknown elements in the input distribution throughout the training phase. This takes place on several levels, employing the algorithms to create the models, much like training machines to learn for themselves. Artificial neurons sometimes referred to as nodes, make up a neural network (figure 4), which is arranged similarly to the way the human brain does. Three layers of these nodes are layered atop one another (Avijeet, 2022):

The input layer

The hidden layer(s)

The output layer



Each node receives information from data in the form of inputs. The node calculates the inputs, multiplies them using random weights, and then adds a bias. To choose which neuron to fire, nonlinear functions—also referred to as activation functions—are used. Convolutional neural networks (CNNs), long short-term memory networks (LSTMs), recurrent neural networks (RNNs), generative adversarial networks (GANs), radial basis function networks (RBFNs), multilayer perceptrons (MLPs), self-organizing maps (SOMs), deep belief networks (DBNs), restricted Boltzmann machines (RBMs), autoencoders, etc. are some of the algorithms used by deep learning models. No network is seen to be flawless, although some algorithms are better adapted to carry out tasks. The Long Short Term Memory Networks (LSTMs) technique will be used in this research project.

2.3.1 Long-Short Term Memory Networks (LSTMs)

Recurrent neural networks (RNNs) with LSTMs can learn and remember long-term dependencies. Long-term memory retention is the default mode of operation (Avijeet, 2022). Over time, LSTMs preserve information. Due to their ability to recall prior inputs, they are helpful in time-series prediction. In LSTMs, four interacting layers connect in a chain-like structure to communicate especially. LSTMs are frequently employed for voice recognition, music creation, and drug research in addition to time-series predictions. An illustration of how LSTMs work is shown in Figure 5 below. The following are the steps for LSTMs.

1. They start by forgetting pointless details from the previous state.

2. After that, they update specific cell-state values.

3. The output of specific cell state components

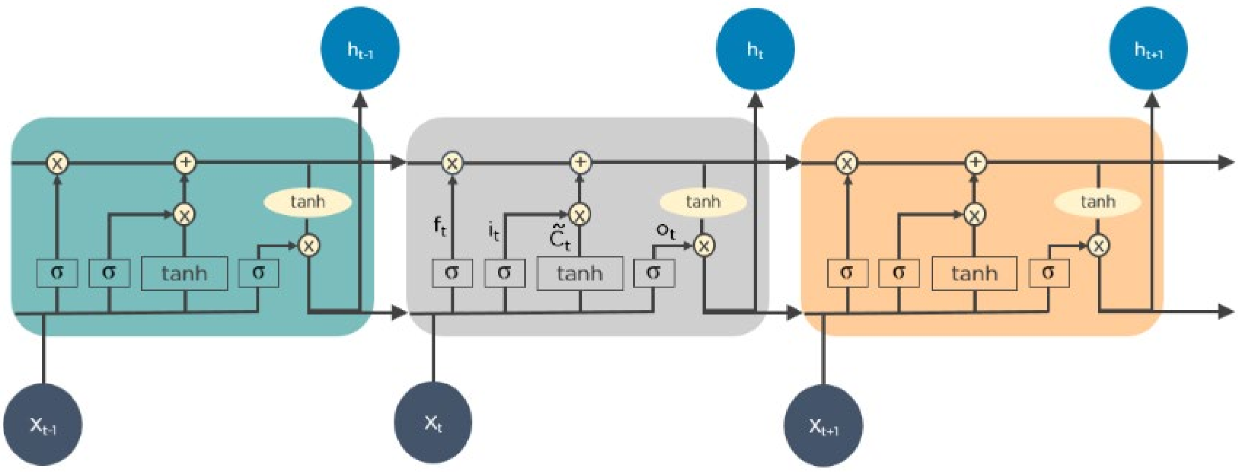


Figure 5: Diagram of how LSTMs operate (Avijeet, 2022)

2.4 Related Literature

To enhance clinical decision-making, Alexander (2017) focused his study on two tasks: integrating natural language processing (NLP) and machine learning to electronic health records (EHRs). The initial aim was to forecast cardiac resynchronization treatment (CRT) outcomes than the current physician recommendations more accurately for the surgery. He trained a supervised classifier to predict CRT results using structured data and NLP features from free-text medical notes. Although the results showed a small improvement over the existing baseline, he was unable to anticipate the outcome of CRT accurately and reliably. These findings restrict the model's clinical utility and support earlier research that failed to identify reliable predictors of CRT response. In his research, Awais (2019) introduced a machine learning-based electronic healthcare forecasting system. The main goal of the thesis is to learn complete patient representations from EHRs. They also provide proof that, unexpectedly, many ML models (such as language processing and time-series prediction) created largely for non-EHR analysis may be merged and modified into a single framework to effectively represent EHR data and forecast patient outcomes (Awais, 2019).

Natural language processing research was presented by Binggui *et al*. (2022). They believe that artificial intelligence (AI) technologies provide a wide range of intelligent applications in a variety of healthcare contexts. From a technological standpoint, they concentrated on feature extraction and modeling for various NLP tasks seen in smart healthcare. The elaboration mainly focuses on typical smart healthcare situations, such as clinical practice, hospital management, personal care, public health, and drug research, in the context of NLP-based smart healthcare apps. They were able to demonstrate the power and potential of NLP for providing smart healthcare through the system they designed. They go on to explore the shortcomings of recent research in the areas of interpretability, human language understanding, and the use of NLP systems for smart healthcare. Finally, they suggested integrating multimodal and longitudinal data, creating end-to-end apps, few-shot learning, and merging several NLP algorithms (Binggui *et al*., 2022). According to Miguel *et al*. (2021), radiology reports are papers that describe and interpret ultrasound pictures. If done correctly, the computerized processing of these texts can aid in the diagnosis of medical practitioners. Regarding the case study of the work, it is impressive that the odd findings were correctly detected because they could have an impact on the patient's health. Focusing on challenging instances can also benefit scholars and medical professionals. In the outpatient healthcare setting Daryl *et al*. (2015) created an android application for an electronic medical record system. The design and implementation of a suggested outpatient management system that enables effective administration of a patient's medical information were given in the study. By fusing a chosen open source EMR system with a proposed Android-based mobile application, they offered a system for scheduling appointments with medical professionals.

With the help of the application, patients and doctors may schedule appointments and use the electronic messaging feature to instantly send reminders as the appointment time approaches. Before the suggested system is implemented in the public health care system, several constraints of the system prototype, such as user authentication and data security, must be effectively addressed (Daryl et al. 2015). Pedro (2019) stated that enterprises should employ use cases for chatbots because of how users are using chat platforms and advances in natural language comprehension. To prioritize the more suitable use cases among a range of prospective use cases for Chatbot implementation, a use case selection method was developed. This technique makes use of the factors acquired. According to Gaurav (2019), free text allows clinicians to record detailed information about patients in narratives and first-person accounts. By developing and analyzing prototype systems for both clinical care and research applications, he illustrates this strategy in his dissertation. He created an interactive platform that allows clinicians to train and create binary NLP models independently for the PTO review procedure notes in the past. The outcomes of the creation and assessment of these prototypes will offer perception into the generalized design of interactive NLP systems for more extensive clinical applications (Gaurav, 2019).

According to Sarmad (2020), a Medicare system that serves as a kind of medical social media is being developed. The article created the Medicare system to utilize strategies including social networking sites, smartphones, wearable technology, and medical equipment to improve the quality of healthcare in Iraq. Through its Android App Interface, the Medicare system is accessible on Android-powered tablets and smartphones. Additionally, electronic devices such as PCs, cellphones, and tablets can use the Medicare web application. Only a few categories of research have been done on a system like the one that is proposed that is aimed at the Iraqi environment, especially those from rural areas who cannot receive health services via the Internet, and the current effort seeks to address this gap (Sarmad, 2020). According to Deepa (2012)'s research, the adoption of smartphones and tablets has inspired a lot of curiosity among healthcare professionals. A growing number of medical institutions have started developing courses for smartphones and tablets. By removing the need to be tied to a workstation, a portable tablet that makes clinical documentation easier can increase the mobility of residents and doctors. A clinical evaluation tool for syncope was created on an iPad to test its usability in this setting given the popularity of Apple's tablet computer. The main goal of the thesis was to create and evaluate a mobile tablet software for clinical evaluation. The app's information is based on clinical practice recommendations. Using organized, prepopulated items and unstructured free-text narratives, the app support clinical evaluation. The study demonstrated that an app with a focus on usability during design and development may be created using evidence (Deepa, 2012). According to Sabina (2013), the use of wireless networks and mobile applications is expanding quickly across a variety of global industries. The researcher has provided a comprehensive architecture for a secure mobile healthcare system in the thesis. The maintenance of patient medical records in a local setting is provided by this application. The Android platform was used to create the mobile application. Because it satisfies crucial security needs, such as integrity, confidentiality, and availability, this solution is sufficiently safe. The researcher proposed that the same mobile application that was put into use may be improved with features for remote patient monitoring for the elderly or disabled. In that circumstance, patient data should be gathered via wireless sensor devices. The patient's mobile phone can then get the data that was collected. After that, a mobile device can send data to a server for a healthcare database (Sabina, 2013).

According to Sumithra *et al*. (2018), during the past few years, there has been a growing understanding of the significance of using Natural Language Processing (NLP) techniques in clinical informatics research, which has resulted in revolutionary advancements. A special emphasis is focused on mental health research, a field that has received little attention from the clinical NLP research community but in which NLP techniques have significant application. Although there have been considerable recent improvements in clinical NLP method development, it was suggested that for the field to grow further, rigorous assessment needs to receive greater attention. According to Sumithra *et al*. (2018), the necessity of using Natural Language Processing (NLP) techniques in clinical informatics research has come to light more and more over the past few years, and this has resulted in revolutionary advancements. Research on mental health is given special attention because it is still largely understudied by the clinical NLP research community even though NLP techniques are very applicable in this field. Although recent developments in clinical NLP technique development have been noteworthy, they also suggested that for it to go further, more emphasis needs to be given to rigorous evaluation. In a paper, Ana (2016) discusses tools to support Process Discovery as well as approaches for the BPM life cycle phases of Process Identification, Process Discovery, and Process Analysis. The present study's findings were helpful for future work on extracting business process models from natural language literature (Ana, 2016).

According to Abu *et al*. (2012), "smartphones" are portable computing devices that can run third-party applications and have extensive mobile communications capabilities. Healthcare professionals are among the constantly expanding demographic of smartphone users. They wanted to categorize smartphone-based healthcare innovations that have been studied in academic literature according to their functions and to provide summaries of each category's publications. To find papers that explored the design, development, testing, or application of smartphone-based software for medical or nursing students, healthcare professionals, or patients, MEDLINE was searched. Many hospitals utilize a variety of paper forms to document their contacts with patients, according to Harrison *et al*. (2017); while these have been quite successful, the world is moving toward digitalization, necessitating the need for paperless medical records. A prototype web application system was developed and assessed for the client using the project management approach Scrum, which is backed by the workload measurement method NASA TLX and the usability evaluation method System Usability Scale (SUS). With the help of this technology, users may easily record measured health data on their smartphones in a very easy manner and monitor changes in their long-term health conditions. According to Nwabueze and Oju (2019), the usage of mobile healthcare applications has emerged as a key area of innovation that may help people manage their daily healthcare needs. In addition to helping people modify their behavior and prevent illness, mobile applications can help improve healthcare delivery, lower costs, and increase effectiveness. The article offers a prototype for an interactive mobile healthcare application that satisfies the demands of both patients and clinicians. For better healthcare services and delivery, the suggested mobile application would offer the best communication to nuseveralyers in the healthcare sector, including patients, physicians, and pharmacists.

2.5 Summary of Literature Review and Research Gap

In the context of processing medical information, this chapter reviewed the literature and methods for deep learning and natural language processing currently in use. Numerous authors suggested other methods, including web-based solutions. Some flaws were found in each of the works that were assessed, and this leaves room for development. Daryl *et al*. (2015), research project, in which they created an Android application for an electronic medical record system in an outpatient healthcare setting, stands out among these techniques. Before the suggested system is implemented in the public health care system, several constraints of the system prototype, such as user authentication and data security, must be effectively addressed (Daryl *et al*. 2015). Natural language processing research was presented by Binggui *et al*. in 2022. They suggested integrating multimodal and longitudinal data, creating end-to-end apps, few-shot learning, and merging several NLP algorithms (Binggui *et al*. 2022).

By incorporating one-time passwords (OTPs) and natural language processing, which enables the system to record physician voice during consultations, convert it to text, and then store it in the medical record database, the proposed system will fill the research gap left by the reviewed studies in terms of user authentication and data security.

3. METHODOLOGY

3.1 Methodology

The new system was created using JavaScript, PHP, and MySQL to analyze deep learning and natural language processing systems in the context of processing medical information. The research project initially looks at the difficulties in Nigeria's current medical information processing system. After that, a system was created that extracts voice data from medical encounters with doctors and converts it to text using Natural Language Processing (NLP) and Deep Learning (DL) techniques. Additionally, this included provisions for the creation of automated systems for processing medical records in hospitals. In the new system, we'll consider employing a hybrid strategy (a password plus an OTP) to secure the processing of medical information. As a result, the application software will be more secure. This research endeavor used the object-oriented analysis and design methodology (OOADM) for software development. A collection of guidelines for system analysis and application design is known as OOADM. The study and design of information systems are done using a formal, rigorous approach. The analysis models are developed further using object-oriented design (OOD) to create implementation requirements. The fundamental distinction between object-oriented analysis and other types of analysis is that with the object-oriented method, requirements are organized around objects, which combine states (data) and actions (processes) modeled after actual items with the system interacting within the real world.

As previously said, Java script and PHP-MySQL will be used to construct the suggested system, and a MySQL database will be used to implement the back end.

3.2 Software / Libraries to Use

The deep learning and natural language processing system was developed in the context of processing medical information using the PHP-MySQL program and Java Script. We will be able to accomplish the goals mentioned by using PHP-MySQL. The user-friendly nature of PHP-MySQL makes it possible to create interfaces that can be changed programmatically. A reliable database that can ensure database integrity, database protection, and support huge databases is MYSQL. Before deciding on PHP and JavaScript, several issues were considered, including deep learning, online database access, data transmission via networks, database security, voice data capture, etc.

3.3 A General Overview of the Processing of Medical Data

Data warehouses (DW) were used by the health sector to integrate data into medical intelligence systems. A registry must be an active instrument utilized frequently by individuals who care for patients inside the clinic for it to be effective. Registries offer quick access to complete, pertinent patient information through printed patient reports when a patient is seeing the doctor. In addition to informing the doctor that the patient has one or more chronic conditions, the report provides a snapshot of the patient's state, saving them time from having to search the patient's medical file for details relevant to the disease. The patient report also includes revisions to the registry's patient data for future data submission. To cut down on the time spent on these duties, most registry programs currently in use nearly entirely rely on information obtained from other electronic systems.

The receptionist determines that the patient is on a register from a number on the patient's card, asks for a printout of the patient report, and attaches it to the patient's folder. In certain offices, the patient record is printed and recorded in the medical record after each visit, and the patient may also be identified in the scheduling software used at the front desk (to be referenced the next time). The doctor plans the service delivery and updates the patient report as necessary. While conducting patient interviews, taking, and documenting vital signs, and placing essential service orders by standing order rules, medical assistants and nurses consult the registry report. The report can be used by doctors for a variety of things, including:

- Patients' diagnosis
- Tell the receptionist when to schedule the next follow-up appointment.
- Keep track of patient information updates for registry entry.

Typically, a central site gathers patient reports for batch entry, though occasionally a care team member does this work. The whole data flow diagram for the current medical information processing is shown in Figure 6.

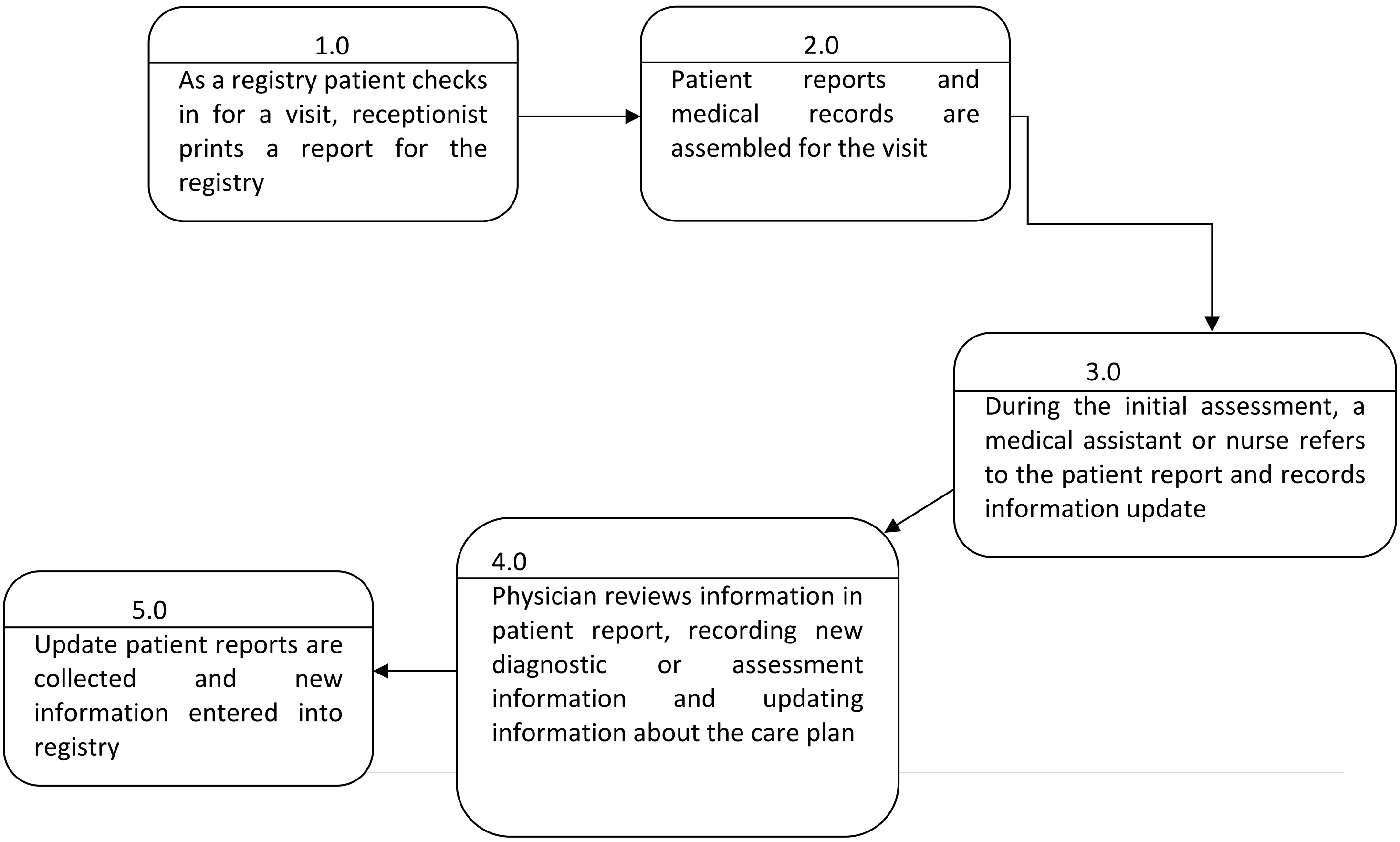


Figure 6 provides a broad overview of the data flow diagram used in the current medical information processing system

3.4 Dataset

Heart diseases, commonly referred to as cardiovascular diseases, will be studied in the dataset used in this study (CVDs). It is recognized as the leading cause of mortality worldwide, accounting for an estimated 17.9 million deaths year, or roughly 32% of all fatalities (kaggle.com/datasets, 2022). Coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other illnesses are among the category of heart and blood vessel disorders known as CVDs. Heart attacks and strokes account for four out of every five CVD deaths, and one-third of these deaths happen before the age of 70. One of the largest heart disease datasets available for research, the dataset has 1191 records and is composed of 10 similar features. The dataset uses the following features.

1. The patient's years of age (Numeric)
2. Gender of the Patient Male as 1 Female as 0 (Nominal)
3. Chest discomfort is divided into four categories: typical, typical angina, non-anginal pain, and asymptomatic.
4. Blood pressure in mm/HG in the resting state (Numerical)
5. The mg/dl of serum cholesterol (Numeric)
6. If the blood sugar level is over 120 mg/dl when fasting, it is represented by 1 as true and 0 as false.
7. The ECG result while at rest is shown by three different numbers. 0: Normal 1: An anomaly in
8. Attained maximum heart rate (Numeric)
9. Exercise-induced angina 0 represents no, 1 represents yes (Nominal)
10. Compared to the resting state, exercise caused ST-depression (Numeric)

Table 1 displays a sample of the dataset. The appendices will include a copy of the entire dataset.

Table 1: Heart Disease Dataset Samples

Age	Sex	Chest pain type	Resting bp s	cholesterol	fasting blood sugar	resting ecg	max heart rate	exercise angina	Oldpeak	Target
40	1	2	140	289	0	0	172	0	0	0
49	0	3	160	180	0	0	156	0	1	1
37	1	2	130	283	0	1	98	0	0	0
48	0	4	138	214	0	0	108	1	1.5	1
54	1	3	150	195	0	0	122	0	0	0
39	1	3	120	339	0	0	170	0	0	0
45	0	2	130	237	0	0	170	0	0	0
54	1	2	110	208	0	0	142	0	0	0
37	1	4	140	207	0	0	130	1	1.5	1
48	0	2	120	284	0	0	120	0	0	0
37	0	3	130	211	0	0	142	0	0	0
58	1	2	136	164	0	1	99	1	2	1
39	1	2	120	204	0	0	145	0	0	0
49	1	4	140	234	0	0	140	1	1	1
42	0	3	115	211	0	1	137	0	0	0
54	0	2	120	273	0	0	150	0	1.5	0
38	1	4	110	196	0	0	166	0	0	1
43	0	2	120	201	0	0	165	0	0	0
60	1	4	100	248	0	0	125	0	1	1
36	1	2	120	267	0	0	160	0	3	1
43	0	1	100	223	0	0	142	0	0	0
44	1	2	120	184	0	0	142	0	1	0
49	0	2	124	201	0	0	164	0	0	0
44	1	2	150	288	0	0	150	1	3	1
40	1	3	130	215	0	0	138	0	0	0
36	1	3	130	209	0	0	178	0	0	0
53	1	4	124	260	0	1	112	1	3	0
52	1	2	120	284	0	0	118	0	0	0
53	0	2	113	468	0	0	127	0	0	0
51	1	2	125	188	0	0	145	0	0	0
53	1	3	145	518	0	0	130	0	0	1
56	1	3	130	167	0	0	114	0	0	0
54	1	4	125	224	0	0	122	0	2	1
41	1	4	130	172	0	1	130	0	2	1
43	0	2	150	186	0	0	154	0	0	0

Source (datasets.kaggle.com, 2022)

3.5 A description of the proposed system for deep learning and natural language processing in the context of processing medical data

Figure 7 illustrates the user interface (UI) and backend of the proposed NLP-driven application. The backend receives text or speech input from the user via the user interface (UI), processes it using NLP models, and then returns the results to the user by offering certain services via the UI. Backend knowledge bases are also necessary for applications that primarily rely on knowledge. Through speech, writing, and other means, the UI enables information interaction, for improving the user experience with intelligent systems and accomplishing smart medical information processing, easily accessible user interfaces are essential. Utilizing NLP techniques, particularly speech recognition and natural language comprehension, one can develop such user interfaces. The fundamental design of NLP-driven applications is depicted in Figure 7.

Speech recognition can be used in the proposed system to take free text notes, which can cut down on the amount of time medical staff spends on the labor-intensive clinical recording. Additionally, the suggested system included clinical decision support (CDS) systems, which can offer doctors recommendations for diagnoses and treatments utilizing deep learning. Heart illness, often known as cardiovascular disease, will be the focus of the CDS (CVDs). The heart illness dataset (hears that log Cleveland and Hungary final) from kaggle.com/datasets will be used for deep learning (2022). between users and intelligent systems. The system was created with the utmost consideration for security, continuity, and accessibility of health information across space and time.

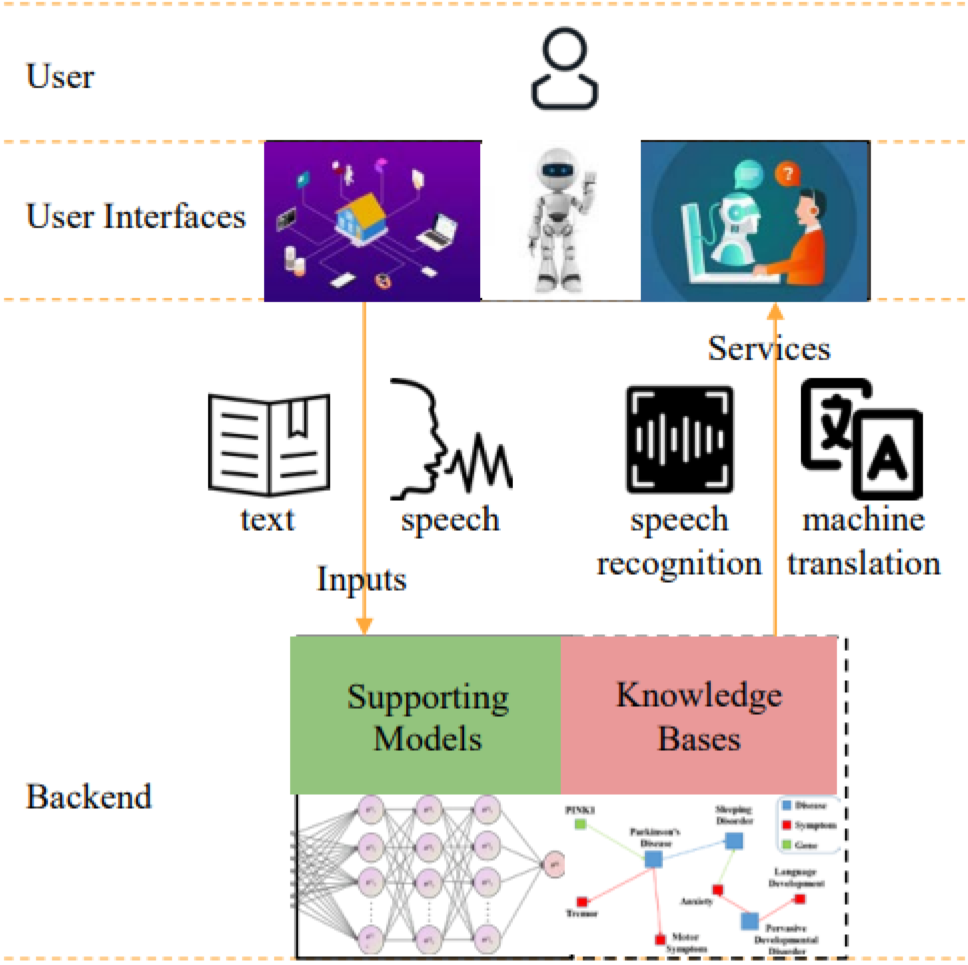


Figure 7 illustrates the fundamental design of NLP-driven apps.

3.5.1 Advantage of the Proposed System

The suggested system will have the following benefits:

1. Voice-to-text conversion that is automatic
2. The system would ensure real-time processing, analyzing, and accessing of data.
3. The system would include a convenient platform for communication between medical professionals and the public.
4. The system would be knowledgeable, trustworthy, adaptive, adaptable, flexible, and agile.
5. The system would be reliable because it guarantees data security and that managers at all organizational levels (operational, tactical, or strategic) make wise, accurate, and timely medical decisions.

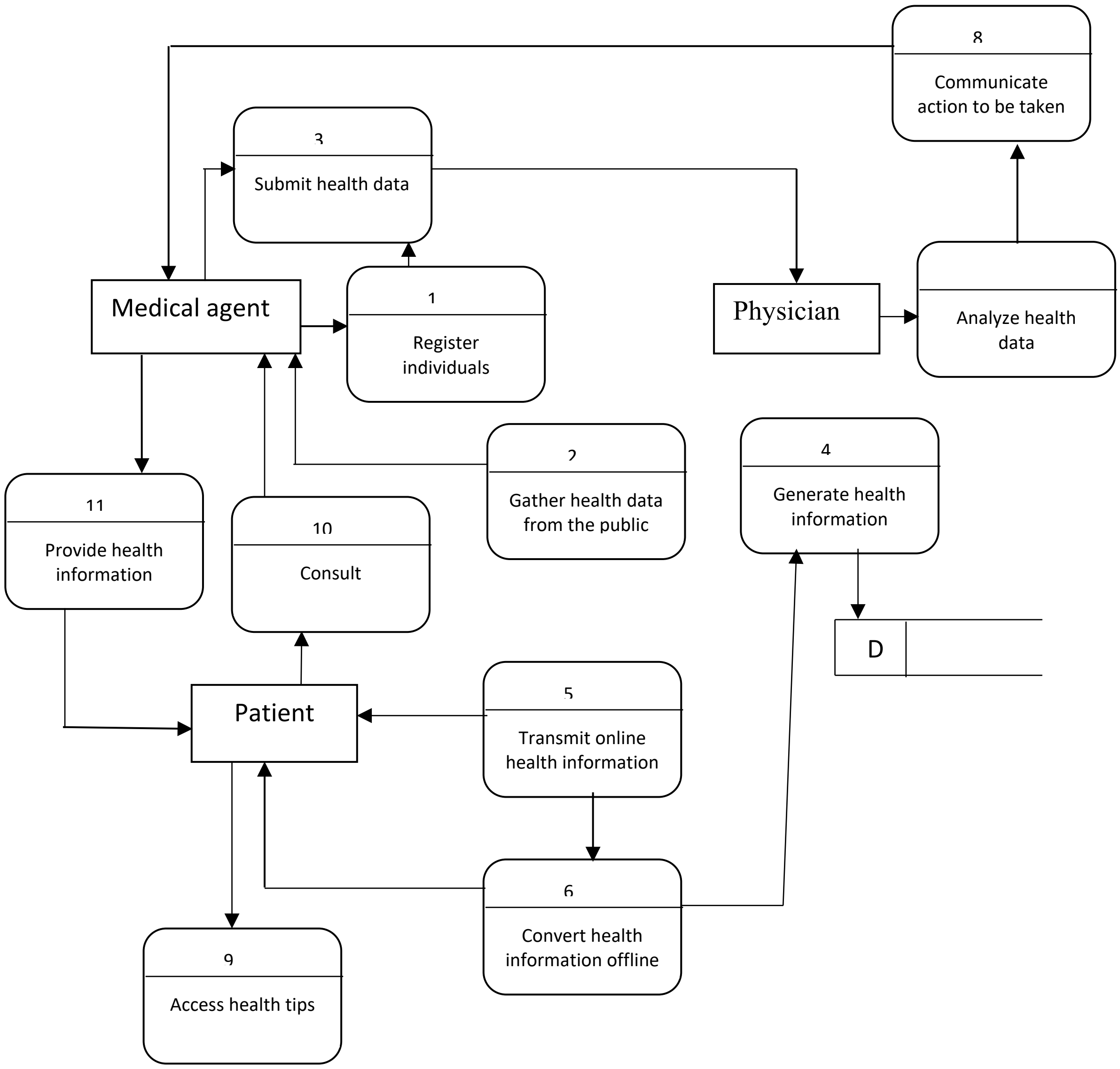


Figure 8: shows the proposed system's data flow diagram.

The health industry oversees the medical data from each data center, as shown in figure 8. The tertiary hospital, clinics, and community healthcare center make up each information integration platform, which enables the multiple levels of medical institutions in a region to share data from data centers and implement two-way referral and medical record lending.

Use Cases Diagram of the Proposed System's

The suggested system is composed of several modules that require access controls. The application of several use cases in the software design was described. Figure 9 below lists the use cases.

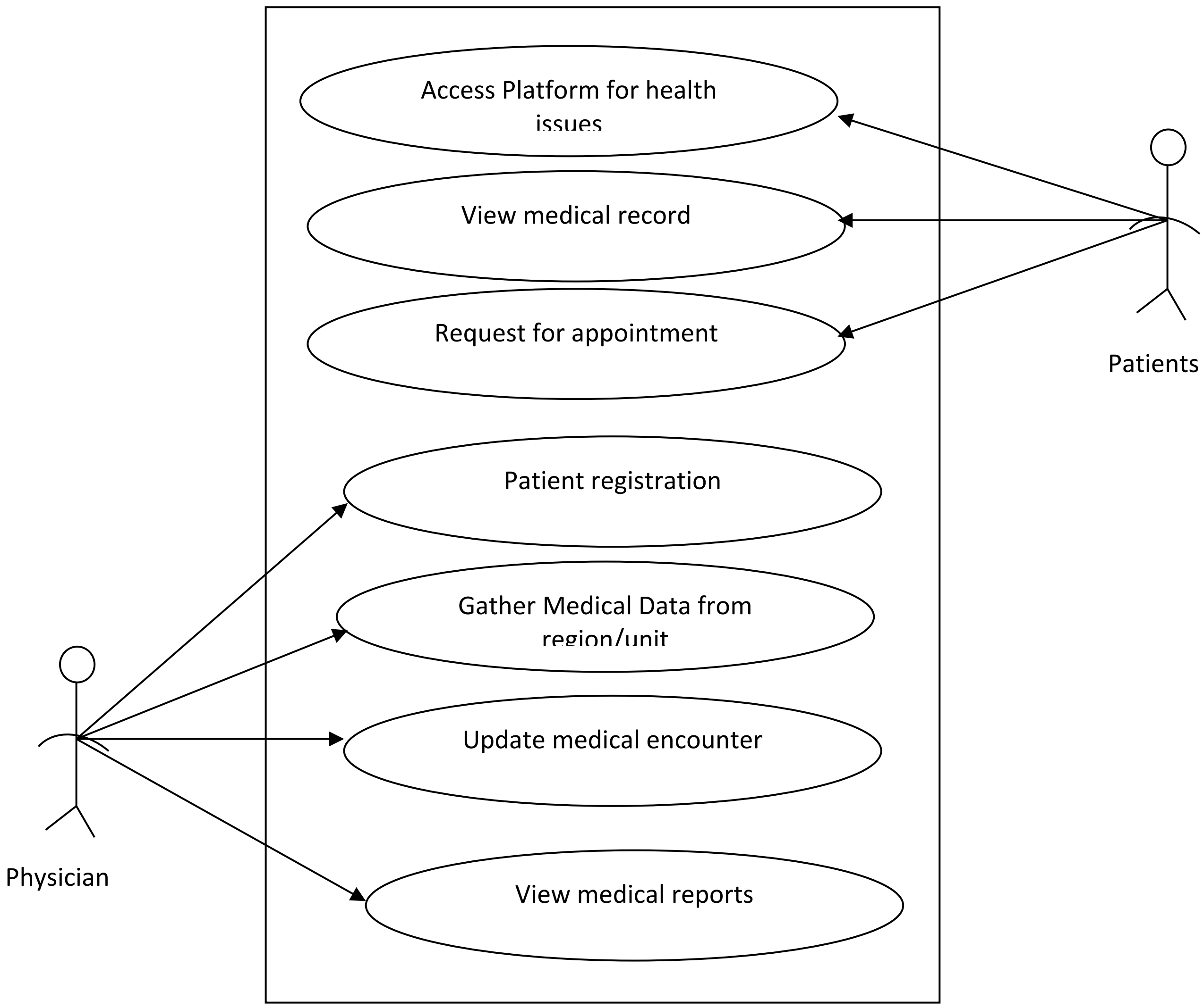


Figure 9: Use case Diagram of the Proposed System

3.5.2 Interaction Diagram

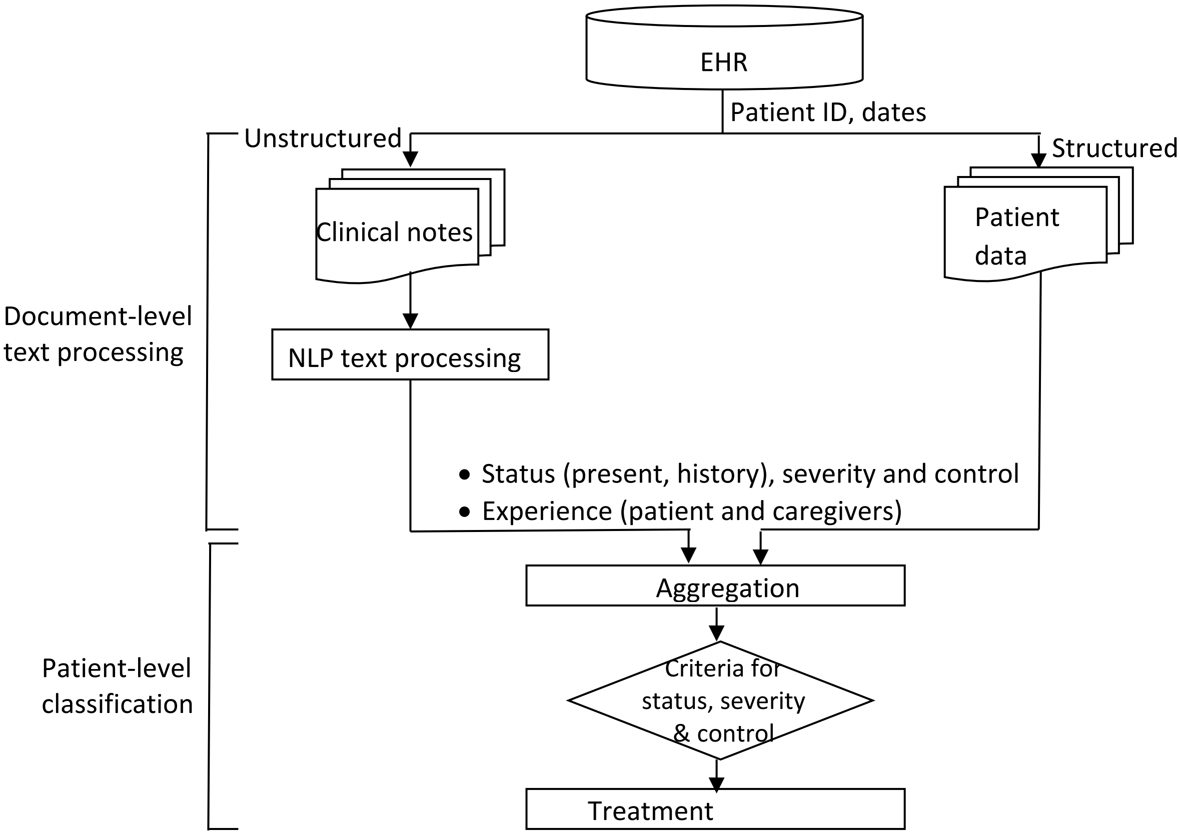
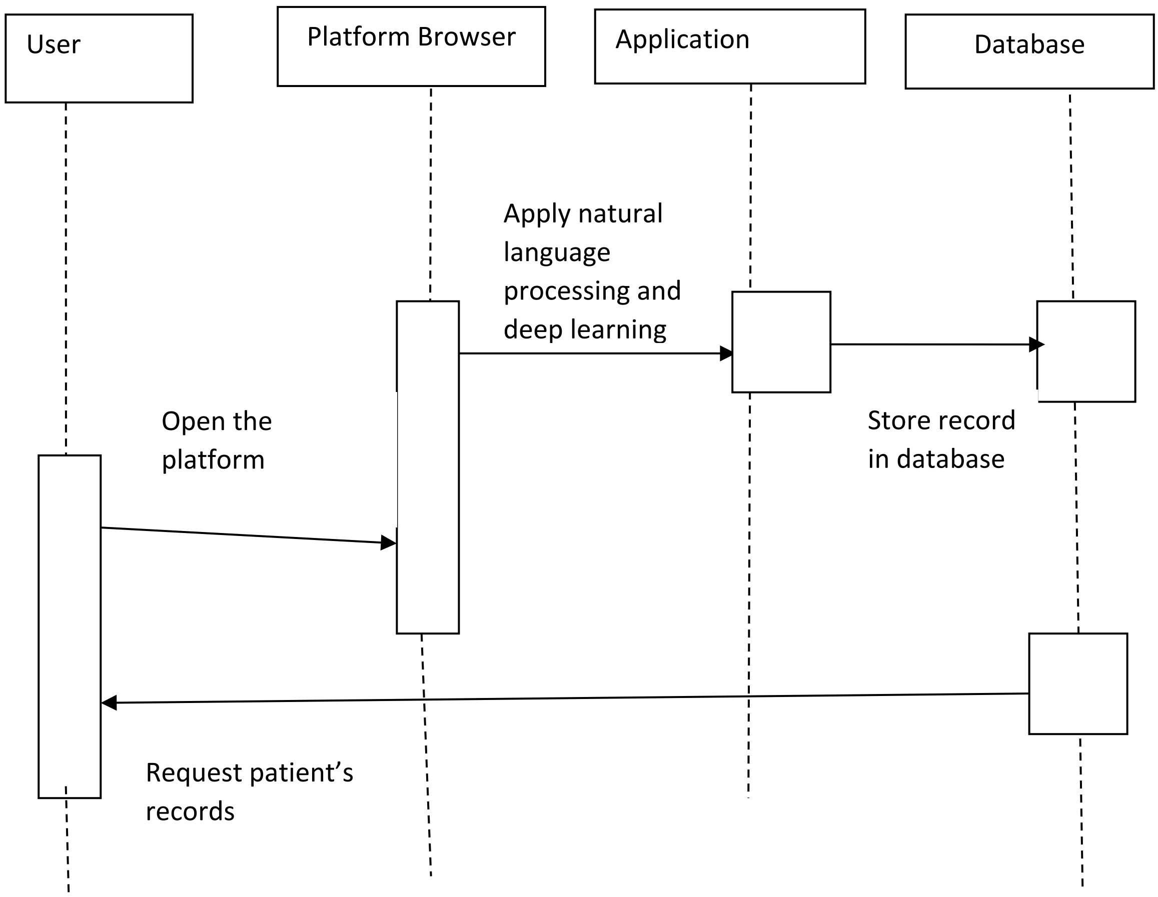


Figure 10: Interaction Diagram of the new System

The text pre-processing and patient classification shown in Figure 10 is based on an interaction diagram or NLP algorithms for document-level text processing and patient classification.

3.5.3 Sequence Diagram

Figure 11's sequence diagram demonstrates the order in which several things interact with one another. It illustrates the kind of objects and classes included in the scenario.



3.6 High-Level Model of the Proposed System.

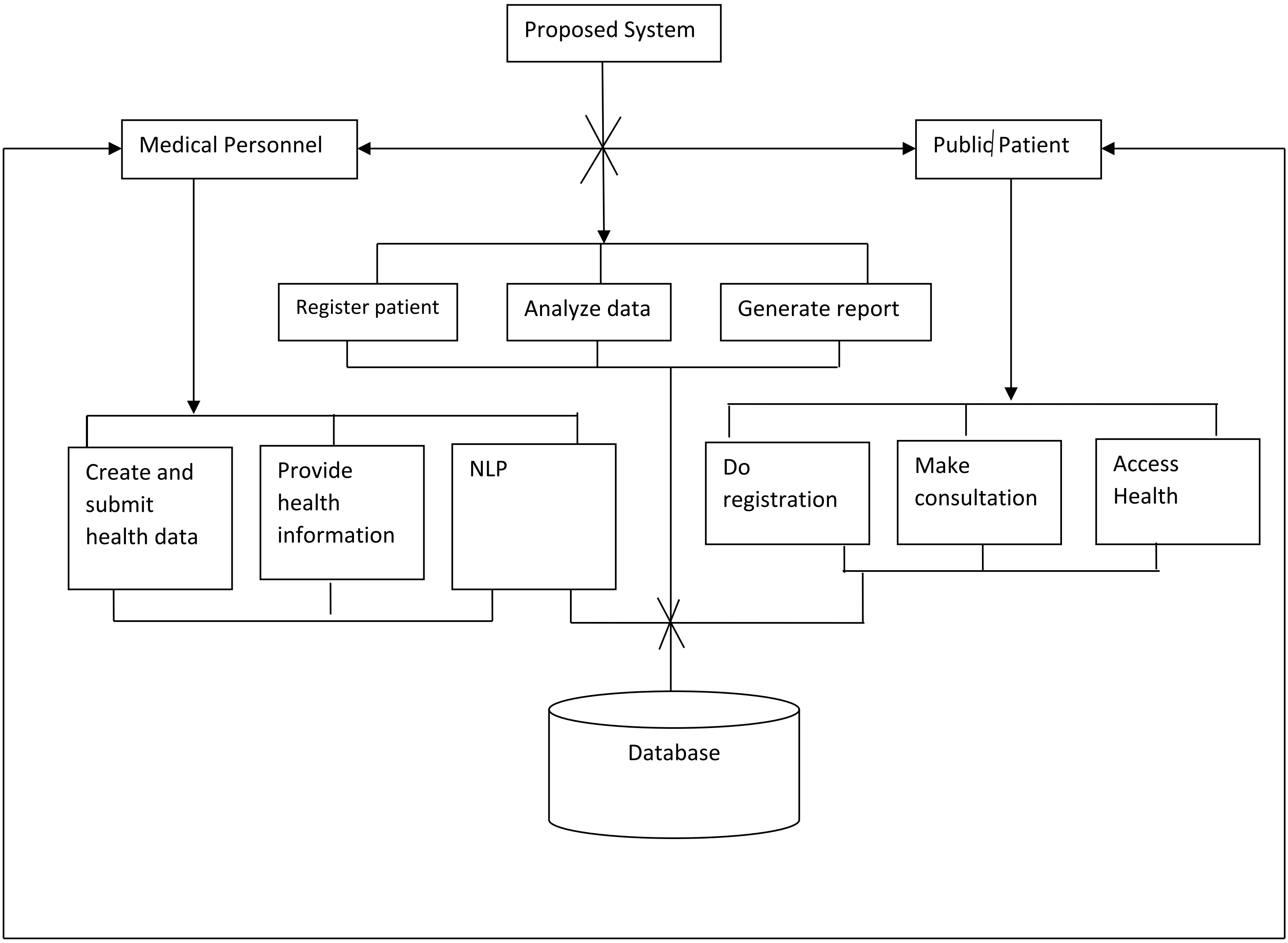


Figure 12: High-Level Model of the Proposed System

Following is a detailed discussion of the high-level model's functions.

1. The proposed system has nine processes, including the creation of profiles, data analysis, report generation, creation and submission of health data, provision of health information and consultation, registration, consultation, and access to health, with three objects to be included: medical personnel, the proposed system, and the public/patient.
2. The user (public/patient) created a page with his or her profile, which includes medical history, after registering with the medical unit.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1 New System Requirements

4.1.1 Hardware Requirements

The following hardware specification were used to implement the analytical, deep learning, and natural language processing systems in the context of processing medical data:

1. 2.4 GHz processor
2. 4 GB RAM; and
3. 1 TB hard drive
4. Printer
5. Internet Modem

4.1.2 Software Requirements

In the context of processing medical information, the software requirements for analysis, deep learning, and natural language processing systems include:

1. WAMP Server,
2. Windows 10,
3. Microsoft Dream Weaver 8,
4. JQuery, Fireworks, and
5. PHP-MySQL

4.2 Program Development

4.2.1 Choice of Programming Environment

Many different programming languages were considered when creating this software. Numerous variables, such as online database access, data transmission via networks, database security, online database retrieval, multiuser network access, online data capture, etc., were considered.

This work made use of PHP-MySQL and JavaScript to accomplish the goals. Additionally, PHP-MySQL allows for the creation of an interface that can be changed programmatically and is very user-friendly. A reliable database that can ensure database integrity, database protection, and support huge databases is MYSQL.

4.2.2 Justification of Language

Java Script and the PHP-MySQL programming language were selected because they offer the benefits of simple development, flexibility, the capacity to offer the developer potential clues, and the ability to construct a graphical user interface.

4.3 Main Menu Design

4.3.1 Control Centre/Main Menu

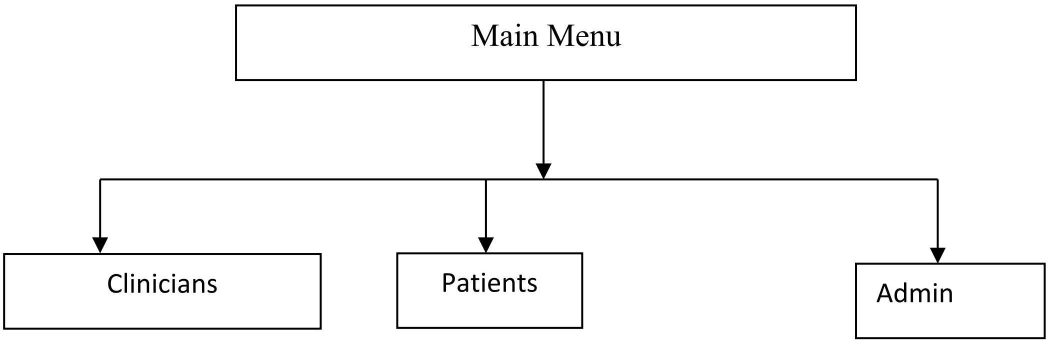


Figure 13: Main menu of the Proposed System

The clinician's module, used by doctors and nurses at various hospitals, the patient's module, from which patients may obtain medical information, and the admin module, used by the software administrator to maintain the database, are all included in the main menu, as seen in figure 13.

4.3.2 Clinician's Sub System

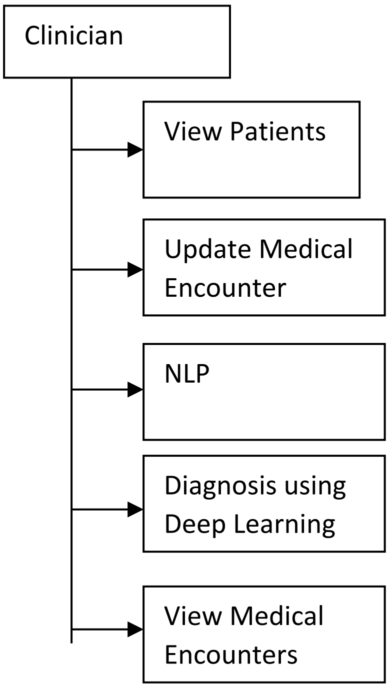


Figure 14: Clinician's subsystem

Clinicians have access to the subsystem depicted in figure 14 and can use it to examine a list of patients, edit a patient's medical encounter, view patients' medical reports, and other tasks.

4.3.3 Patient's Sub System

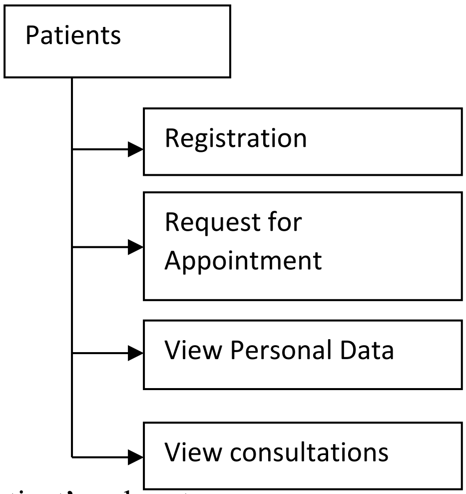


Figure 15: Patient's subsystem

Figure 15 depicts the patient's subsystem and the tasks that can be completed on the newly created system. The patient has access to his or her personal information, medical history, online consultation with a doctor, and appointment requests.

4.3.4 Admin Sub System

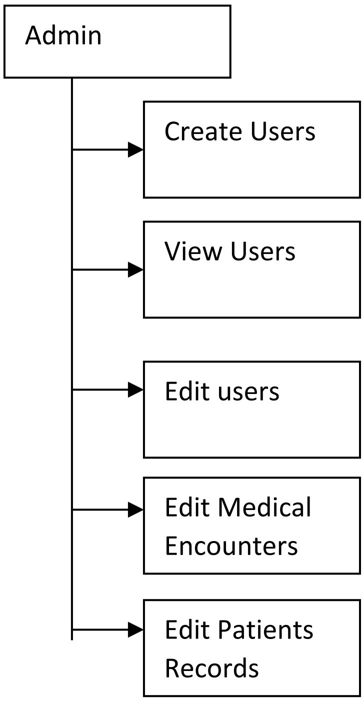


Figure 16: Admin subsystem

Administrators can do tasks like adding users and modifying database information in the admin subsystem, as seen in figure 16.

4.4 Input / Output Format

The following forms show the format of input and output forms as designed in the proposed system:

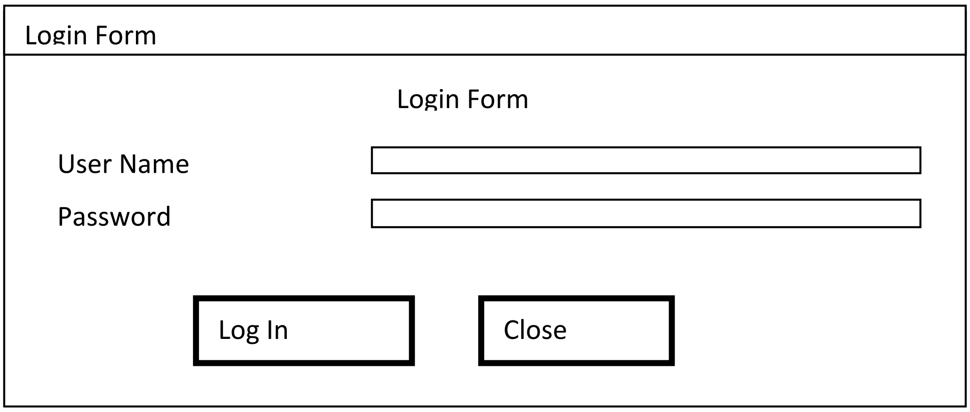


Figure 17: Login Form

The platform user's login information, which includes the username and password, is shown in Figure 17. After entering the specification, pressing the login button verifies the information before launching the user onto the designated subsystem.

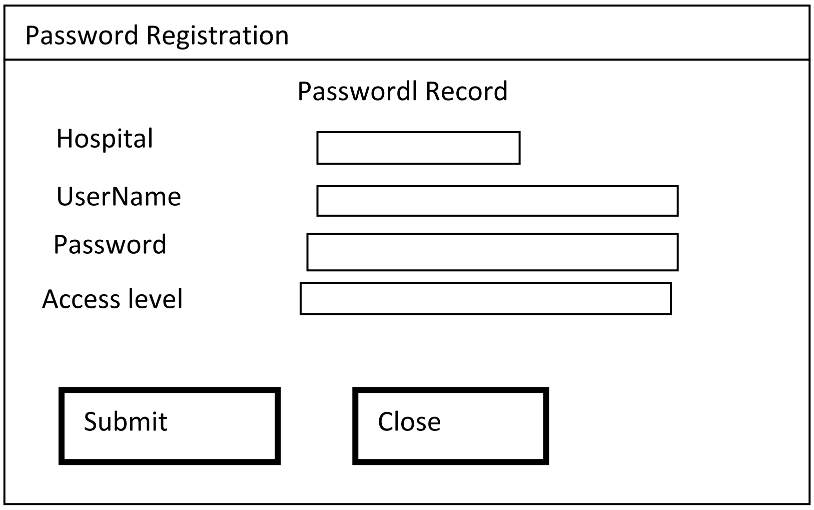


Figure 18: Password Registration Form

To add a new hospital entry to the database, use Figure 18. It creates a hospital access password and registers the hospital.

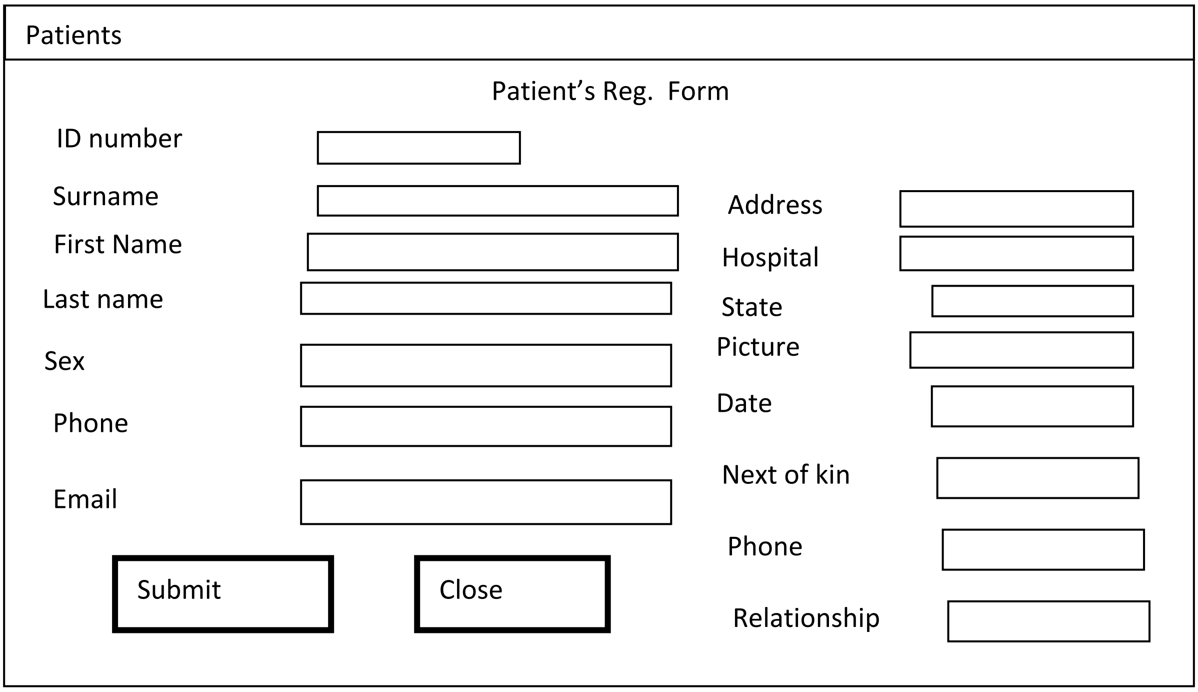


Figure 19: Patient’s Registration Form

The form used in the hospital to enter patient information is depicted in Figure 19. It records the patient's information as well as that of the hospital where the registration took place.

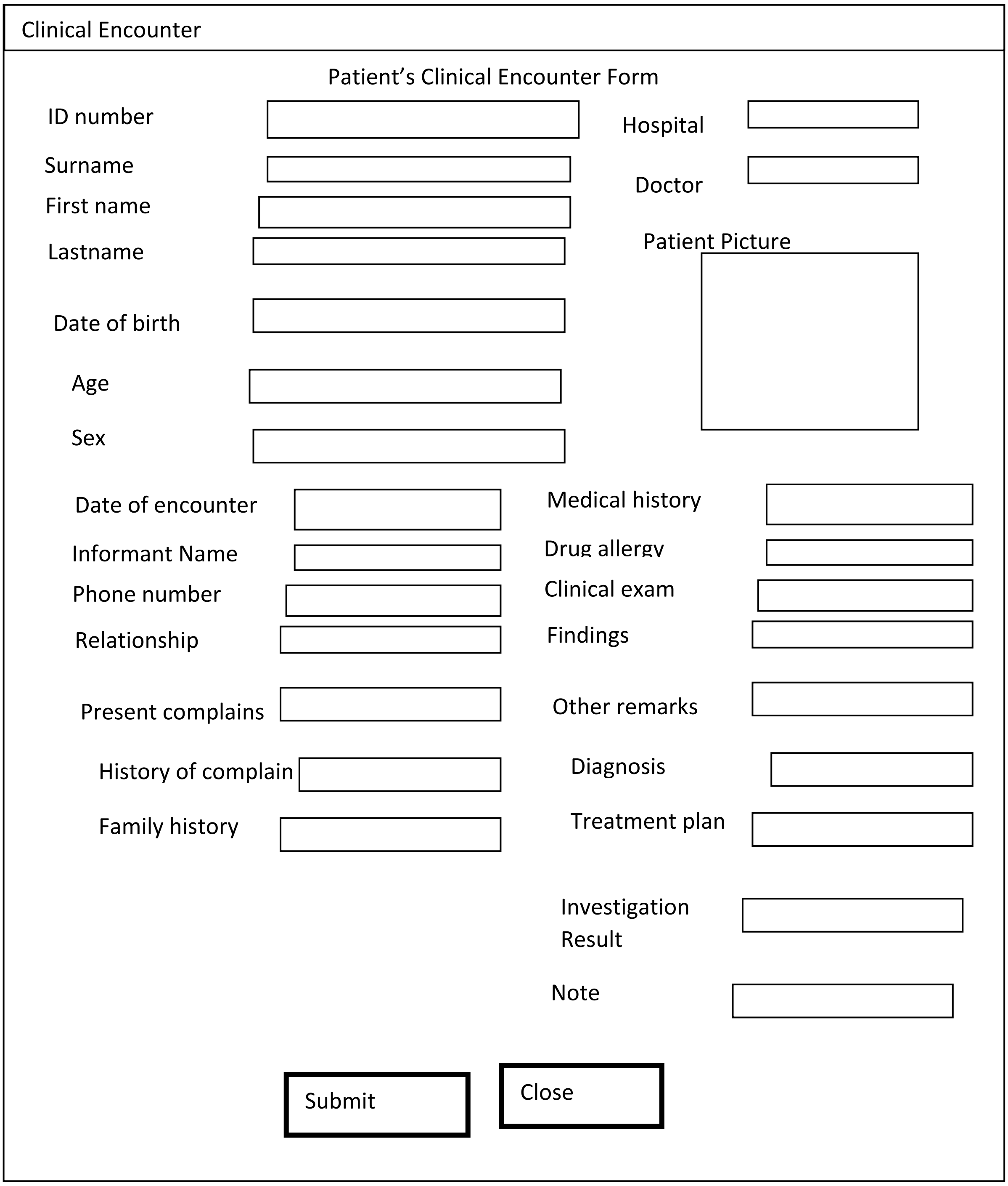


Figure 20: Patient clinical encounter Form

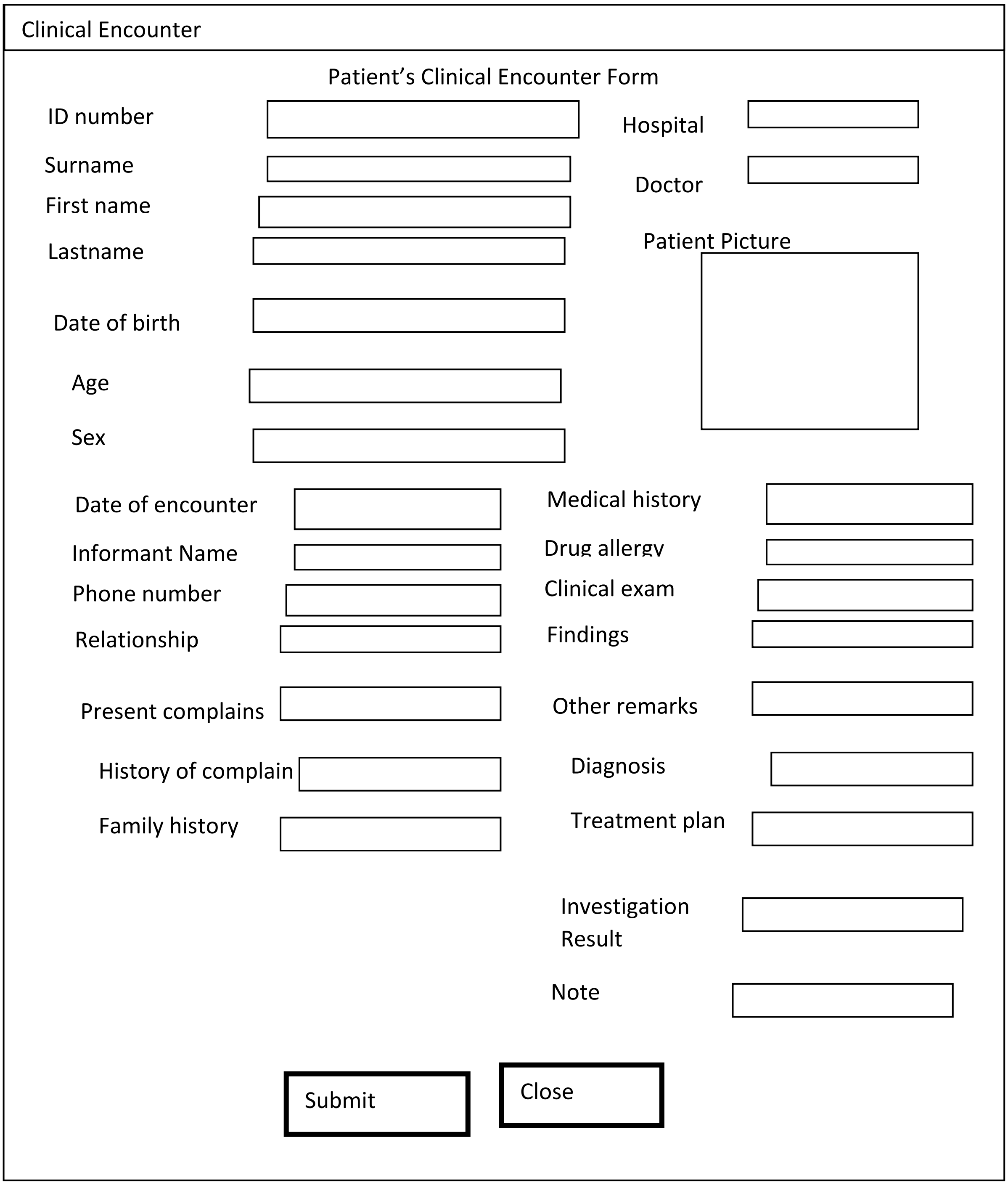


Figure 20: Patient clinical encounter Form

The form used in the hospital to enter patient information is depicted in Figure 19. It records the patient's information as well as that of the hospital where the registration took place.

The form shown in Figure 20 is used to document a patient's clinical interaction with a physician. Using deep learning and natural language processing, this form can record voice and transform it into text.

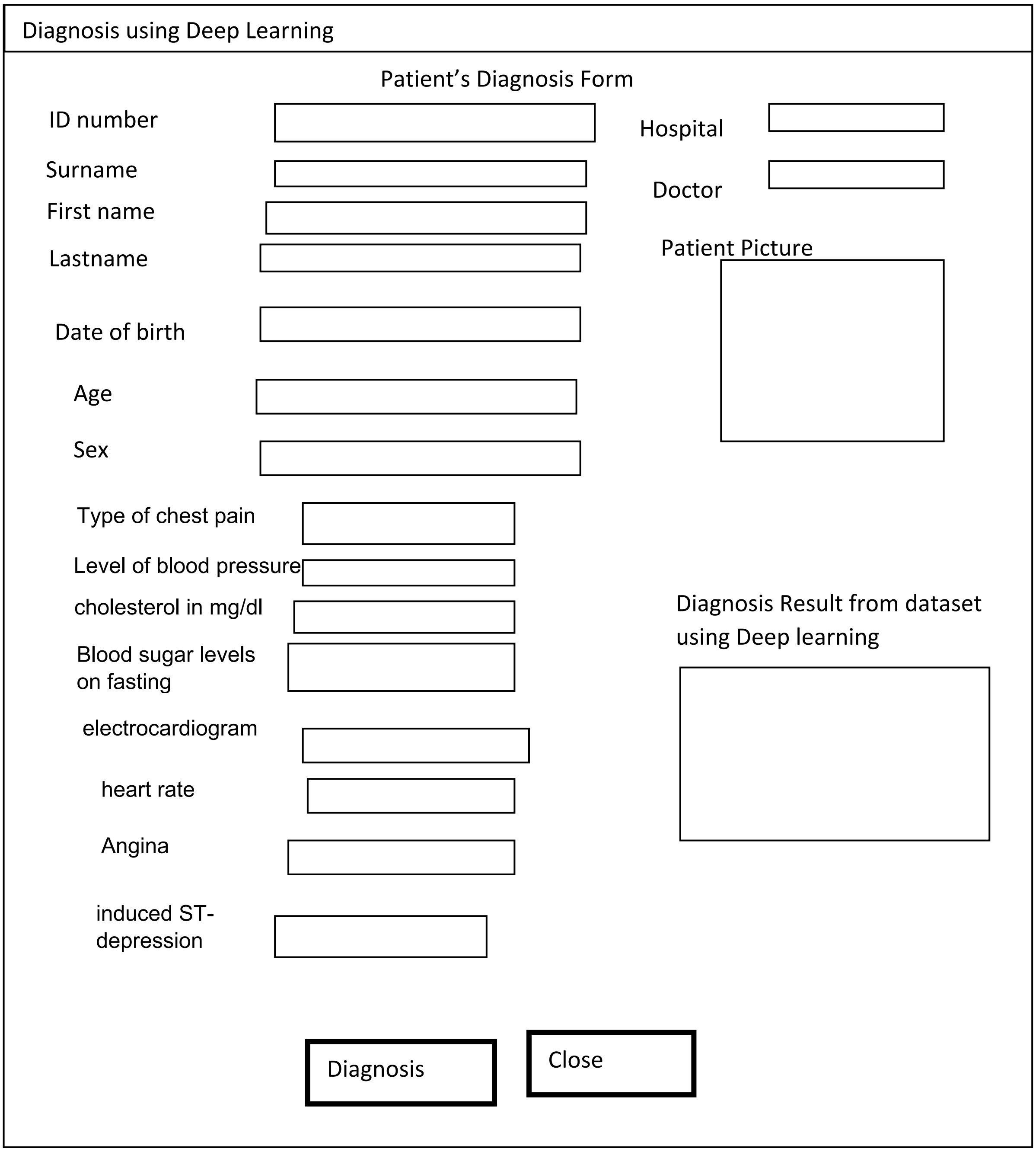


Figure 21: Patient diagnosis Form

The form for diagnosis using deep learning on the dataset to ascertain the results of the parameters gathered is shown in Figure 21.

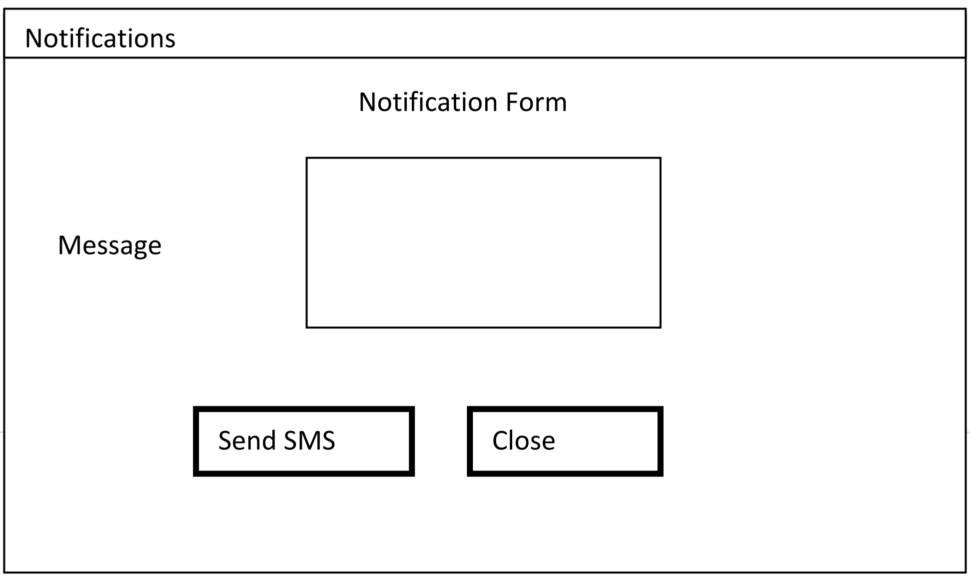


Figure 22: Notification Form

The form used to convey medical information to patients via SMS gateway is shown in Figure 22.

Patient's Registered Report

Surname	Lastname	First name	Address	Email	Phone	Hospital	Patient ID

Figure 23: Patient's File

Figure 23 depicts the patient registry, which contains information about each hospital patient.

Tool for Database Development (4.5)

The following are some of the database development tools used for this thesis:

Wamp Server 2.

Sqlyog Community, Third

Structured query language, or SDL (SQL)

The tables and fields in the proposed system's design were organized using these tools.

4.5. Database Structure and Design

MySQL Database was used in the database design for the Tertiary Unified Medical Records for Integrated Care proposal. The purpose of table structures was to store data for public health information. The definitions of tables 2 through 4 that were designed in the database are as follows.

Table 2: **Users Login Table Structure**

Field	Data Type	Size	Description
Username	Varchar	20	The Users name.
Password	Varchar	12	The password of the user
Level	Varchar	5	Access privileges
Hospital	Varchar	100	The hospital assigned the password to

Table 3: Patient Table Structure

Field	Data Type	Size	Description
Surname	Varchar	15	The surname of the patient
First name	Varchar	15	First name of the patient
Lastname	Varchar	15	Last name of the patient
Sex	Varchar	8	Gender of the patient
Phone	Varchar	12	Phone number of the patient
Email	Varchar	40	Email of the patient
Address	Varchar	150	Address of the patient
Hospital	Varchar	150	Hospital where the patient is treated
State	Varchar	50	State where the hospital is located
Id	Varchar	20	Patient identification number
Pic	Varchar	200	Picture of the patient
Age	Int	3	Age of patient
Dob	Date	8	Date of birth of patient
Marital status	Varchar	20	Marital Status
Religion	Varchar	50	Religion of worship
Lga	Varchar	50	Local Government area
Nationality	Varchar	50	Nationality
Occupation	Varchar	100	Occupation of patient
Kin	Varchar	50	Next of kin
Kin address	Varchar	150	Address of next of kin
Kin phone	Varchar	30	Phone number of next of kin
Kin relationship	Varchar	50	Relationship to the patient

Table 4: **Clinical Table Structure**

Field	**Data Type**	**Size**	**Description**
Tdate	Date	8	Date clinical record was recorded
Patientid	varchar(20)	20	Patient identification number
Informant	Varchar	50	The person that brought the patient to the hospital
Informantphone	Varchar	11	Phone number of the person that brought the patient to the hospital
Informantrelationship	Varchar	20	Relationship to the person that brought the patient to the hospital
Complain	Varchar	500	Patients complaint / symptoms
Historycomplain	Varchar	500	How long the symptoms has been observed
Familyhistory	Varchar	1000	Patient's family medical history
Medicalhistory	Varchar	1000	Patient's medical history
Allergy	Varchar	500	Allergic to any drug
Examination	Varchar	1000	A medical examination conducted by doctor
Findings	Varchar	1000	Medical findings from the examinations
Comments	Varchar	1000	Doctors remark
Diagnosis	Varchar	1000	Recommended diagnosis to be carried out
Treatmentplan	Varchar	1000	Treatments recommended by a doctor
Results	Varchar	1000	Lab result of the diagnosis
Followupnote	Varchar	1000	Medical foll0w up for the patient
Hospital	Varchar	100	Hospital where the patient consulted
Doctor	Varchar	50	The doctor that attended to the patient

4.6 How to Install the Software

To install the software developed on the system, follow the procedure below.

i. Install Micro media Dreamweaver 8 on the computer

ii. Install wampserver2.5-Apache-2.4.9-Mysql-5.6.17-php5.5.12-32bon the computer

iii. Click Start Button on the desktop

iv. Select program

v. Click Windows explorer

vi. Click Drive D:

vii. Select the folder “medicalinformation”

viii. Click Edit

ix. Click Copy

x. Select drive C:

xi. Select Wamp

xii. Select www

xiii. Click paste to Copy the Folder “medicalinformation” from your flash drive to the folder C:\wamp\www

xiv. Open internet explorer

xv. Type http://localhost/medicalinformation

xvi. Select options from the menu

5. TEST RESULTS AND EVALUATION

5.1 Test Result

This section describes the application's implementation and testing for problems and non-functional qualities including processing medical data securely, quickly, and robustly. To guarantee that the specified objectives have been accomplished correctly and the research questions have been addressed to produce a high-quality, user-friendly application, the test is simply the execution of the implemented application using sample data. The various modules of the application were tested, and some of those components are shown in figures 24, 25, 26, 27, 29, and 30.

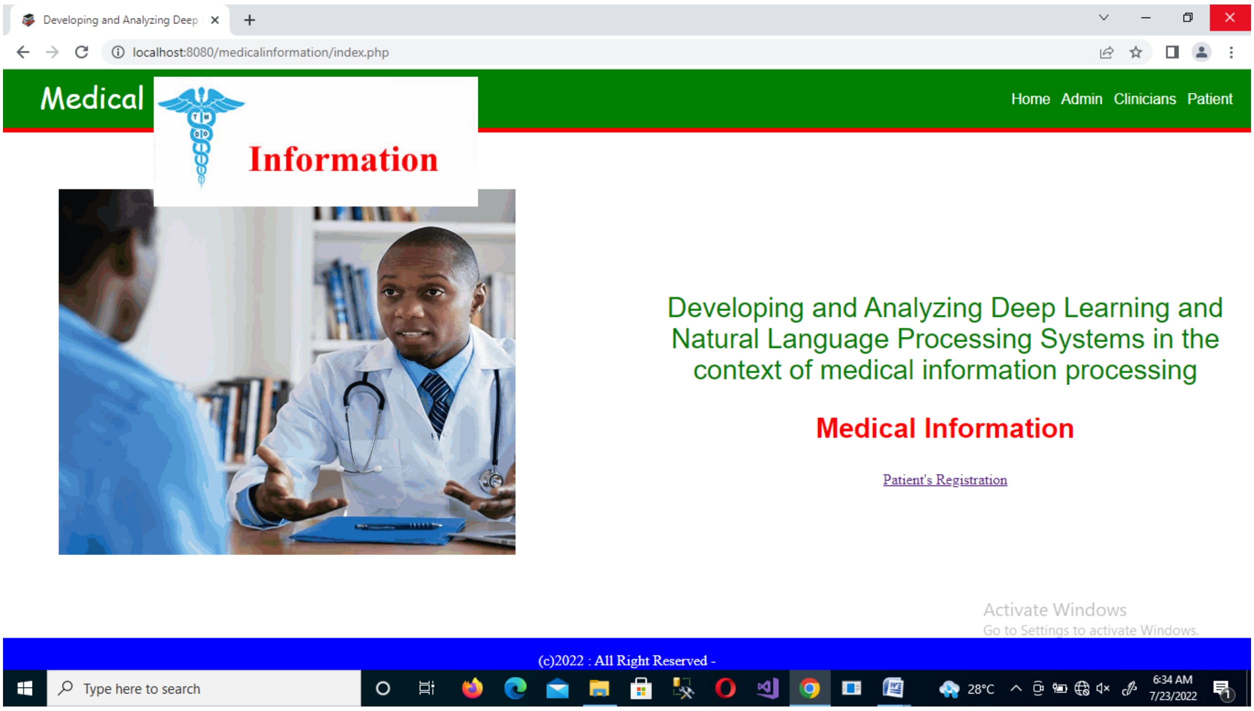


Figure 24: The main page for medical information processing

The application index page, shown in Figure 24, is the first screen you see when the Medical Information Processing platform is launched. Users can use this interface to access the login form.

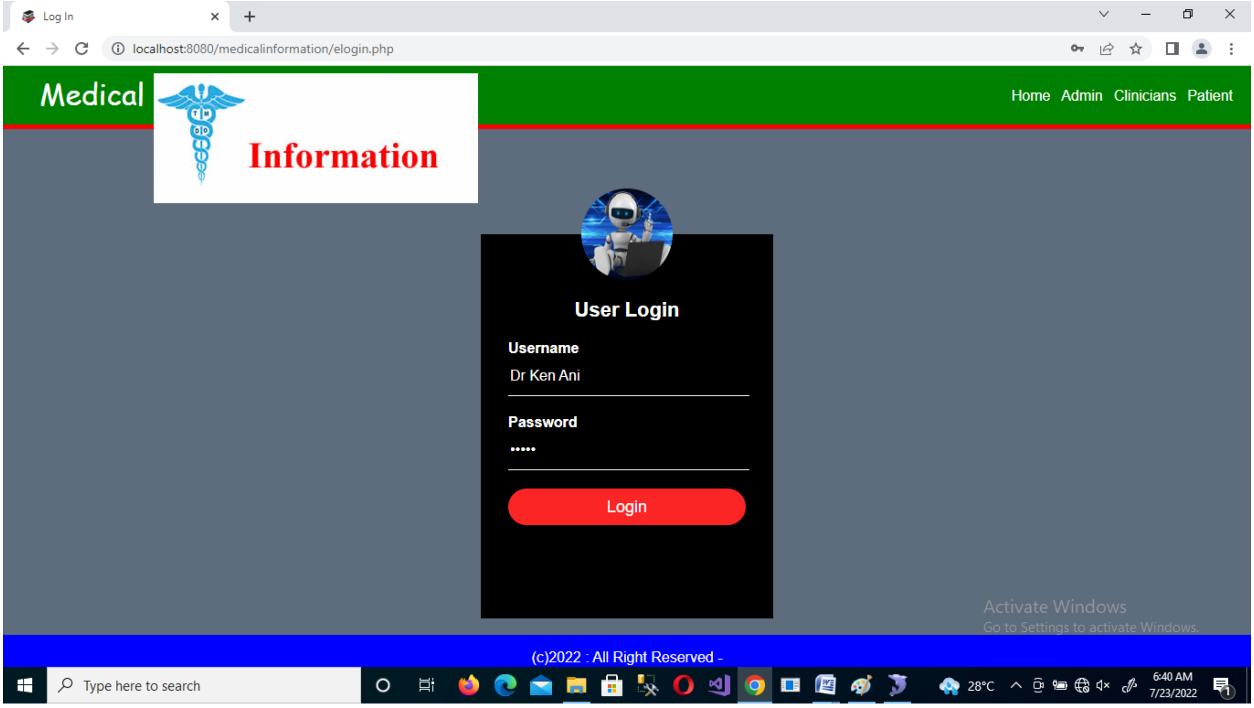


Figure 25: Login page

The initial security authentication for each user on the platform is shown in Figure 25. The username and password must be supplied by the user for authentication. At this point, the system checks the password entered by the user and disables access if the password doesn't pass the authentication test.

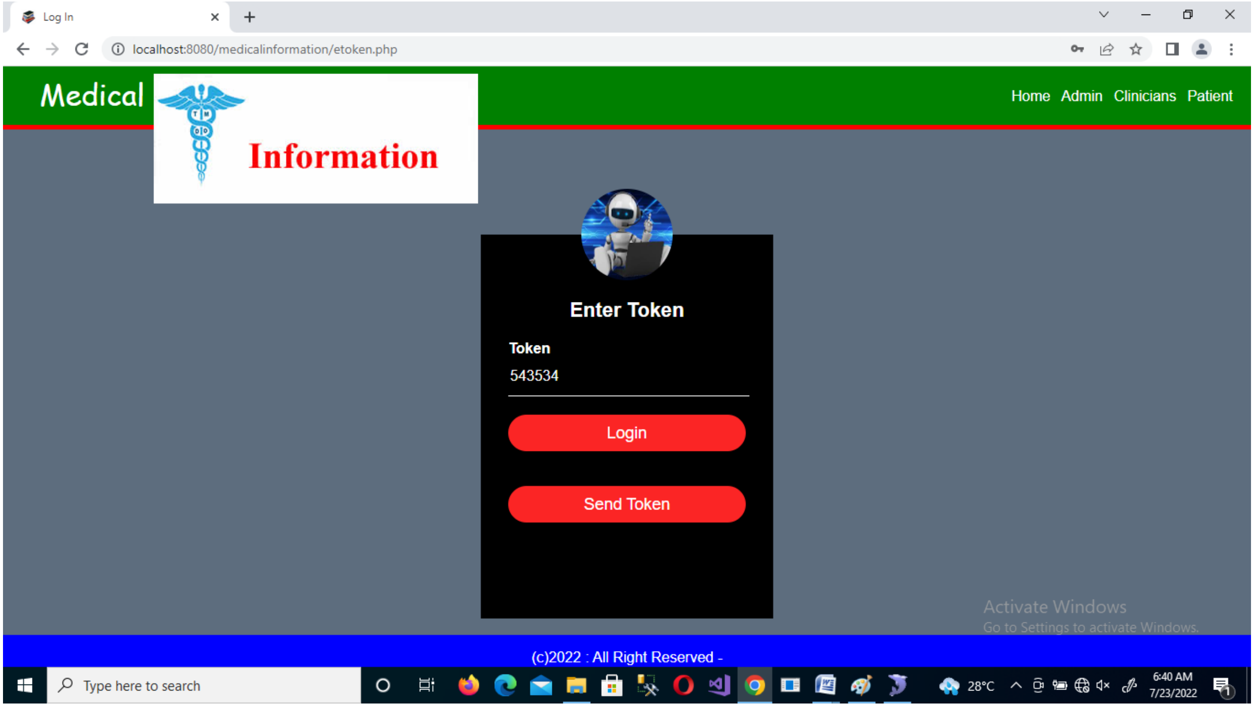


Figure 26: Token Verification page

The second security authentication for each user on the platform is shown in Figure 26. To authenticate the token, the user must provide the SMS receipt they got on their phone. The user will be able to choose options from the application menu after the token has been verified.

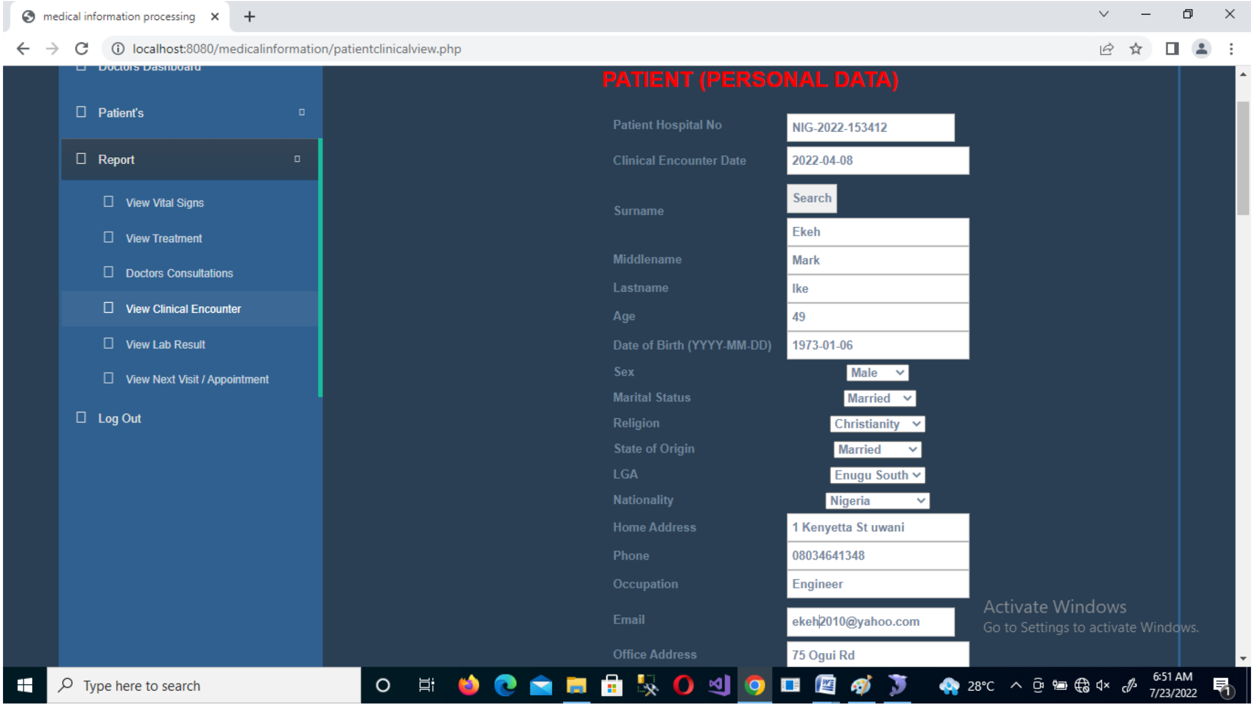


Figure 27: Patient Clinical Encounter / Voice data capturing page 1 (patient personal data)

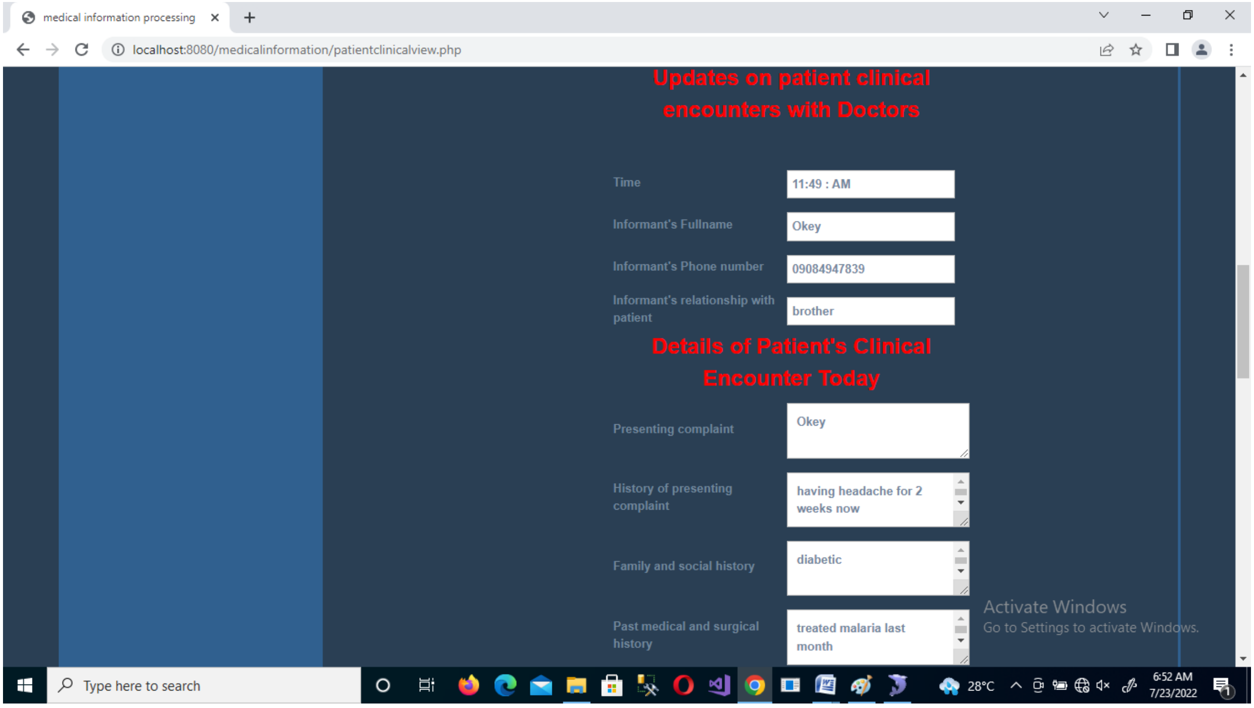


Figure 28: Patient Clinical Encounter / Voice data capturing page 3 (patient clinical data)

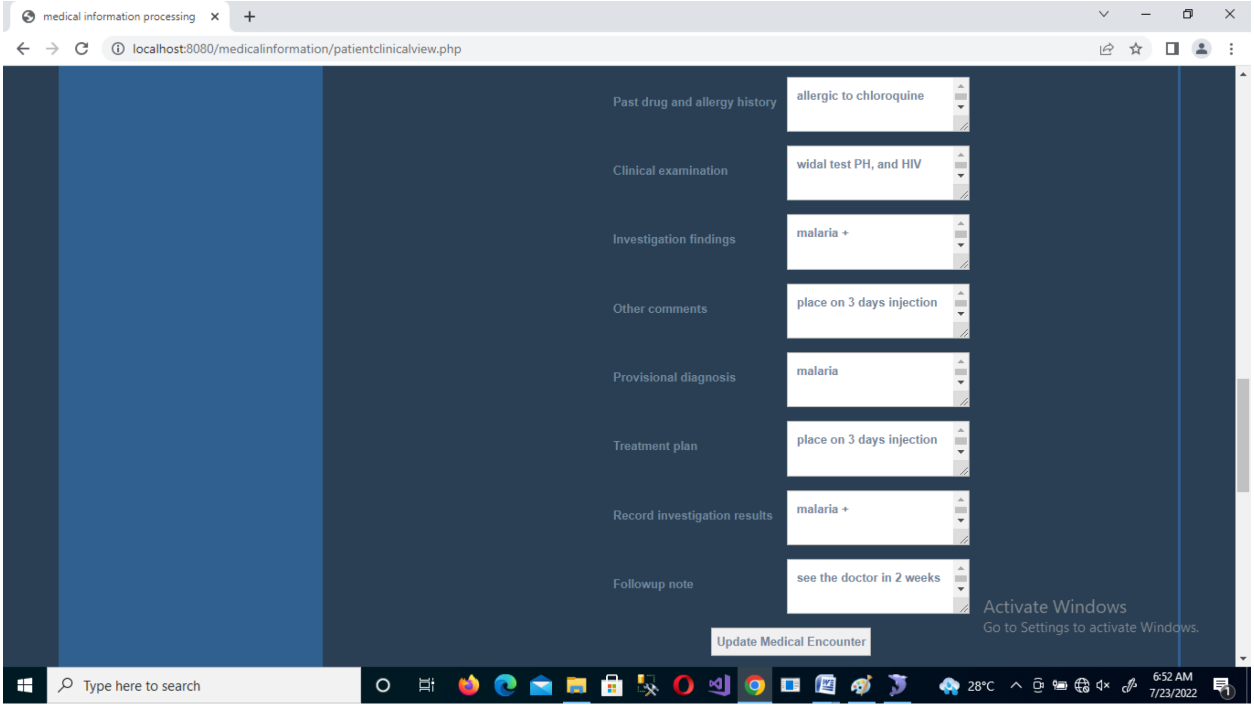


Figure 29: Patient Clinical Encounter / Voice data capturing page 3 (patient clinical data continues)

Forms shown in Figures 27, 28, and 19 are utilized during patient-physician interactions. Natural language processing can be used to convert voice data into text utilizing this form's voice data capture feature. Any comments made by the clinician when questioning the patient will be noted on the form and eventually saved in the database.

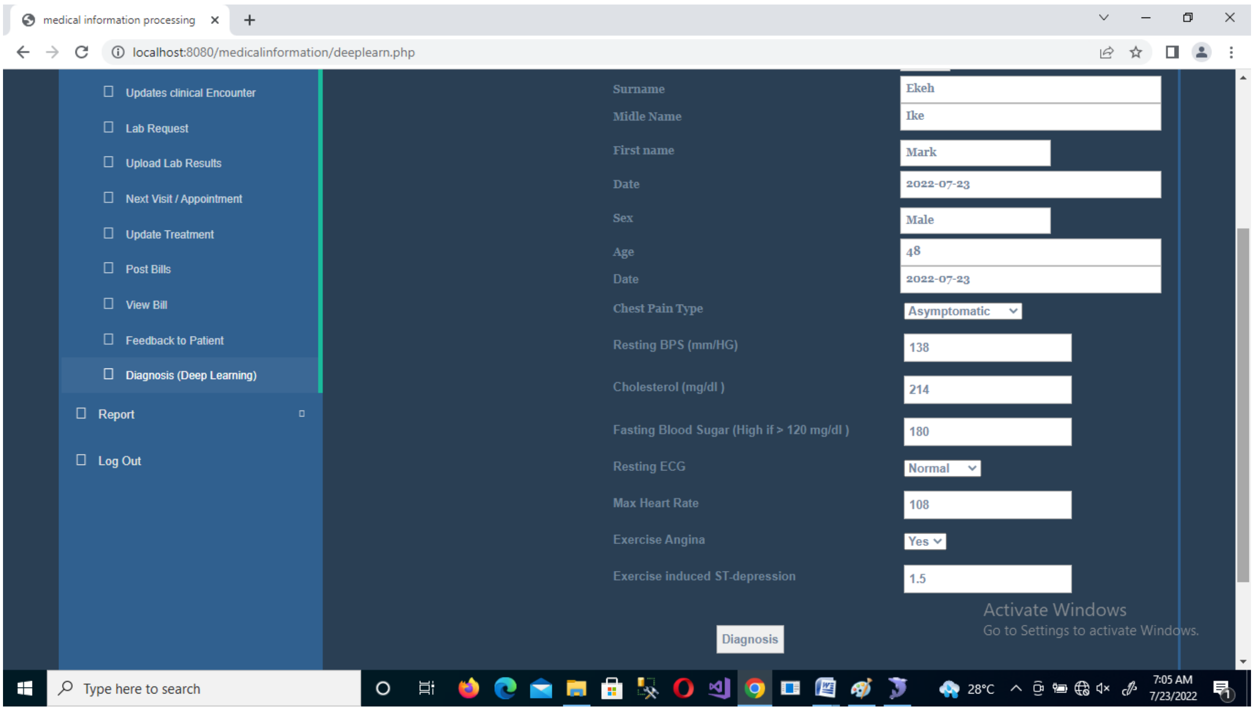


Figure 30: Patient heart disease diagnosis using deep learning on heart disease dataset

To determine whether a patient has a heart-related ailment or not, deep learning is applied to the heart disease dataset using the manner shown in Figure 30.

5.2 Evaluation

The software's performance was evaluated by classifying patient medical data gathered for a heart disease test quickly and accurately using deep learning. One of the largest heart disease datasets available for research purposes, the training data is derived from a dataset with 1191 records and 10 common attributes. The dataset uses the following features.

The collection and production of a large amount of usable data by the medical information processing system allow for the measurement of numerous key performance indicators (KPIs) that support management in making decisions.

Table 5: Confusion Matrix

Observed

		True	**False**
	True	TP	FP
Predicted	**False**	FN	TN

$$AC = \frac{a+d}{a+b+c+d} \quad (1)$$

a = True Positive

b = False Positive

c = False Negative

d = True Negative

Thirty tests for heart disease were conducted as part of the testing to determine whether the system could correctly detect and categorize the patient as having heart disease or not using the data from the dataset and deep learning. The analysis of deep learning and natural language processing systems in the context of processing medical information is graded according to performance in Table 6.

Confusion matrix applied to test dataset in Table 6

Observed

		True	False
	True	12	1
Predicted	**False**	0	17

According to Table 6, 12 of the 30 tests carried out using a deep learning system were True Positives and correctly predicted to have heart disease. 17 were identified as True Negatives, and it was properly predicted that they did not have heart disease. 1 displayed the incorrect categorization and was, therefore, a False Negative. As shown in equation 1, a model of performance measures may be generated from the confusion matrix, demonstrating the system's correctness.

With the numbers substituted, we obtain AC = (12+17) / (12+17+0+1) AC = 0.967, or 96.7 percent accuracy in predicting the result of the diagnostic of heart disease (see table 7).

Table 7: **Performance Results Obtained**

Technique Applied	**Accuracy in classifying the Tickets**
Deep Learning Algorithm	96.7%

6. DISCUSSION AND CONCLUSION

6.1 Discussion

The goal of this research was to develop and analyze deep learning and natural language processing in the context of processing medical information. From the research done, it has been shown that the issues with medical information processing are more related to slow information retrieval and slow medical diagnosis and documentation. Therefore, the research project modeled a system that can convert voice to text during a medical contact using natural language processing. This was made possible by the software's ability to convert speech data from clinicians' observations made during a patient's medical examination into text using natural language processing. As a result, patient medical information is recorded more quickly and without delay. Additionally, the implementation of medical diagnostic for heart illness utilizing deep learning and heart disease dataset. The results of the implementation demonstrate that a patient can be classified as having a heart-related disease or not with an accuracy of 96.7%.

6.2 Conclusion

Providing high-quality healthcare is a challenging endeavor that is heavily reliant on patient data and medical expertise. As far as possible, decisions regarding a patient's care should be based on data from study rather than just clinical judgment and experience. Deep learning and natural language processing play a significant role in the analysis of medical data. It enables doctors to provide timely, high-quality care to their patients. Medical datasets provide historical data on health problems, diagnostic criteria, and treatment results. This will help the doctors learn from the many thousands of records in the collection. Learning cannot be done manually but using a deep learning algorithm can make learning more efficient, accurate, and useful in clinical settings. Therefore, the system created by this research will help doctors handle medical information so they can provide patients with the high-quality, fast, and accurate care they need.

7. LIMITATIONS

7.1 Shortcoming

Development of software requires high level of programming skills. At first, the researcher would have loved to use some python libraries in developing this work but for the time constraints and the need to finish this dissertation in a recorded time. Nevertheless, other programming languages like php mysql, php -ml and java script were explored for quicker output and conclusion of the dissertation in a recorded time.

PHP-ML is a library developed to handle Machine learning tasks using PHP, and this library includes Machine learning algorithms as well as data processing APIs that can handle data cleanups and feature extractions. Php-ml library includes majority of basic algorithms such as classification, sentiment analysis, neural networks, etc.

Also, one of the limitations of this work is that the researcher went for a cheaper software application because he wouldn't be able to afford cost associated with complex hardware and software development platforms, and only require simple predictions and data analytics that can use php-machine learning to solve the purpose.

8. REFERENCES

Abu, S. M. M., Illhoi, Y., and Lincoln, S. (2012*). A Systematic Review of Healthcare Applications for Smartphones*. http://www.biomedcentral.com/1472-6947/12/67

Alexander, W. F. (2017). Improving Clinical Decision Making with Natural Language Processing and Machine Learning. *Computer Science and Engineering at the Massachusetts Institute of Technology*. 13: 45

Allen, J. (1988) Natural language understanding, Benjamin/Cummings Publishing. Pg. 12: 24 – 35

Ana, C., Almeida, B. (2016). A systematic literature review on Natural Language Processing in Business Process Identification and Modeling. *Computer Science*. 24: 45 - 35

Avijeet, B. (2022). Top 10 Deep Learning Algorithms You Should Know in 2022. https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm

Awais, A. (2019). *Predicting clinical outcomes via machine learning on electronic health records. Halmstad University Dissertations no*. 58 ISBN 978-91-88749-24-6 (printed) ISBN 978-91-88749-25-3 (pdf) Publisher: Halmstad University Press

Bahja, M., Hammad, R. (2018). A user-centric design and pedagogical-based approach for mobile learning. *International Conference on Mobile, Hybrid, and On-line Learning*, pp. 100-105

Bengio, Y., Ducharme, R., Vincent, P. and Jauvin, C. (2013). A neural probabilistic language model. *Journal of Machine Learning Research*, 3 (4),1137–1155

Benotti, L., Martínez, M., Schapachnik, F. (2014). *Engaging high school students using chatbots*. In: Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education, pp. 63-68.

Berry, M. A., and Linoff, G. (2012). Data Mining Techniques: for Marketing, Sales, and Customer support. John Willy & Sons Inc, New York, USA.

Binggui, Z., Guanghua, Y., Zheng, S., and Shaodan, M. (2022). Natural Language Processing for Smart Healthcare. Department of Electrical and Computer Engineering, University of Macau, Macao 999078, China. 11: 24 - 35

Boonstra, A., Broekhuis, M. (2017). University of Groningen, Groningen, The Netherlands: Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions; 13: 29 - 30

Bresnick, J. (2020). "What Is the Role of Natural Language Processing in Healthcare?" Health IT Analytics (2020) Retrieved from https://healthitanalytics.com/features/what-is-the-role-of-natural-language-processing-in-healthcare. 23

Bresnick, J. V, (2020). Uses EHRs, and Natural Language Processing to Spot Suicide Risks. Available from: https://ehrintelligence.com/news/va-usesehrs-natural-language-processing-to spot-suicide-risks

Daryl, A., Bulou, G., Nicholas, R., and Rohitash, C. (2015). *Development of an Android Application for an Electronic Medical Record System in an Outpatient Environment for Healthcare in Fiji*. Technical Report, Aicrg, Software Foundation, Fiji.

Deepa A. M. (2012). A Mobile Tablet App for Clinical Evaluation and Medical Education: Development and Usability Evaluation. McMaster University, Hamilton, Ontario

Esteva, A. (2019). A guide to deep learning in healthcare. *Nature Med*. 25 (2019):24-29

Fayyad, U., Piatetsky, G., Smyth, P. (2016). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*. 19: 24 - 40

Gaurav, T. (2019). *Interactive Natural Language Processing for Clinical Text*. School of Computing and Information. The University of Pittsburgh. 3: 5 -8

Gharehchopogh, F., Khalifelu, Z. (2017). Analysis and evaluation of unstructured data: Text mining versus natural language processing. 5Th International Conference on Application of Information and Communication Technologies (AICT). 9: 12 - 34

Ghosh, S. (2009). Application of natural language processing (NLP) techniques in E-governance. *E-Government Development and Diffusion*, pp. 122-132.

Han, J. and Kamber, M. (2016). *Data Mining: Concepts and Techniques*, 2nd edition Amsterdam: Morgan Kaufmann

Hand, D., Heikki, M., Smyth, P. (2011). Principles of Data Mining. Cambridge. The MIT Press. 10: 34 - 39

Harrison, S., Martin, T., Huseyin, D., Nan, J. (2017). Digitizing a Medical Clerking System with Multimodal Interaction Support. *Research gate Conference Paper*. 5: 10 - 29

Hirschberg, J. and Manning, C. (2015). *Advances in natural language processing Science*, 3,(6), 261-266

Juhn, Y., & Liu, H. (2020). Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *Journal of Allergy Clinical Immunology*, 1(2),463-469.

9. APPENDIX

9.1 Appendix A: Ethics Form

Ethical clearance for research and innovation projects

Project status

Status

Actions

Date	Who	Action	Comments
10:54:00 08 July 2022	Shakeel Ahmad	Supervisor approved	
18:13:00 07 July 2022	Kingsley Chijioke Ogam- Okafor	Principal investigator submitted	

Get Help

Ethics release checklist (ERC)

Project details

Proiect name:	Developing and Analyzing Deep Learning and Natural Language Processing Systems to process medical information processing.
Principal investigator:	Kingsley Chijioke Ogam-Okafor
Faculty:	Faculty of Business, Law and Digital Technologies
	Postgraduate
Level:	Applied AI & Data Science
Course:	COM726
Unit code: Supervisor	Shakeel Ahmad
name:	

Checklist

Question	Yes	No
Q1. Will the project involve human participants other than the investigator(s)?	○	◉
Q1a. Will the project involve **vulnerable participants** such as children, young people, disabled people, the elderly, people with declared mental health issues, prisoners, people in health or social care settings, addicts, or those with learning difficulties or cognitive impairment either contacted directly or via a **gatekeeper** (for example a professional who runs an organisation through which participants are accessed; a service provider; a care-giver; a relative or a guardian)?	○	◉
Q1b. Will the project involve the use of **control groups** or the **use of deception**?	○	◉
Q1c. Will the project involve any **risk to the participants' health** (e.g. intrusive intervention such as the administration of drugs or other substances, or vigorous physical exercise), or involve psychological stress, anxiety, humiliation, physical pain or discomfort to the investigator(s) and/or the participants?	○	◉
Q1d. Will the project involve **financial inducement** offered to participants other than reasonable expenses and compensation for time?	○	◉
Q1e. Will the project be carried out by individuals unconnected with the University but who wish to use staff and/or students of the University as participants?	○	◉
Q2. Will the project involve sensitive materials or topics that might be considered offensive, distressing, politically or socially sensitive, deeply personal or in breach of the law (for example criminal activities, sexual behaviour, ethnic status, personal appearance, experience of violence, addiction, religion, or financial circumstances)?	○	◉
Q3. Will the project have detrimental impact on the environment, habitat or species?	○	◉
Q4. Will the project involve living animal subjects?	○	◉
Q5. Will the project involve the development for export of 'controlled' goods regulated by the Export Control Organisation (ECO)? (This specifically means military goods, so called dual-use goods (which are civilian goods but with a potential military use or application), products used for torture and repression, radioactive sources.) Further information from the Export Control Organisation	○	◉
Q6. Does your research involve: the storage of records on a computer, electronic transmissions, or visits to websites, which are associated with terrorist or extreme groups or other security sensitive material? Further information from the Information Commissioners Office	○	◉

Declarations

I/we, the investigator(s), confirm that:

- [x] The information contained in this checklist is correct.

- [x] I/we have assessed the ethical considerations in relation to the project in line with the University Ethics Policy.

- [x] I/we understand that the ethical considerations of the project will need to be re-assessed if there are any changes to it.

- [x] I/we will endeavor to preserve the reputation of the University and protect the health and safety of all those involved when conducting this research/enterprise project.

- [x] If personal data is to be collected as part of my project, I confirm that my project and I, as Principal Investigator, will adhere to the General Data Protection Regulation (GDPR) and the Data Protection Act 2018. I also confirm that I will seek advice on the DPA, as necessary, by referring to the Information Commissioner's Office further guidance on DPA and/or by contacting information.rights@solent.ac.uk. By Personal data, I understand any data that I will collect as part of my project that can identify an individual, whether in personal or family life, business or profession.

- [x] I/we have read the prevent agenda.