ORAL CANCER DETECTION

Shashikala S V[1], Noor Ayesha S[2], Sahana L M[3], Shankreppa Handargal[4], Shifaali[5]

Department of Computer Science and Engineering,

B.G.S Institute of Technology, B.G Nagara-571448, Nagamangala (Tq), Mandya, India[1-5]

Abstract : Oral cancer is a major global health issue accounting for 177,384 deaths in 2018 and it is most prevalent in low- and middle-income countries. Enabling automation in the identification of potentially malignant and malignant lesions in the oral cavity would potentially lead to low-cost and early diagnosis of the disease. Building a large library of well annotated oral lesions is key. As part of the MoMoSA (MobileMouth Screening Anywhere) project, images are currently in the process of being gathered from clinical experts from across the world, who have been provided with an annotation tool to produce rich labels. A novel strategy to combine bounding box annotations from multiple clinicians is provided in this project. Further to this, deep neutral networks were used to build automated systems, in which complex patterns were derived for tackling this difficult task. Using the initial data gathered in this study, two deep learning-based computer vision approaches were assessed for the automated detection and classification with ResNet-101 and object detection with the Faster R-CNN. We create a methodology to extract features from image and implement genetic algorithm and apriori algorithm for association mining to get accuracy more in results.

Key Words: Malignant lesion, Machine Learning

1.INTRODUCTION

The term cancer is used generically for more than 100 different diseases including malignant tumours of different sites (such as breast, cervix, prostate, stomach, colon/rectum, lung, mouth, leukaemia, sarcoma of bone, Hodgkin disease, and non-Hodgkin lymphoma). Common to all forms of the disease is the failure of the mechanisms that regulate normal cell growth, proliferation and cell death. Ultimately, there is progression of the resulting tomur from mild to severe abnormality, with invasion of neighbouring tissues and, eventually, spread to other areas of the body. The primary risk factor for developing oral cancer is tobacco use. Smoking cigarettes, cigars, and pipes all increase risk of oral cancer. Smokeless tobacco, often called "dip" or "chew," also heighten the risk. Alcohol consumption is another habit that is strongly associated with the development of oral cancer. This research uses data mining technology such as classification, association rules to identify potential oral cancer patients. Apriori & Genetic algorithm is the originality algorithm of Boolean association rules of mining frequent item sets. The datamining methods and techniques will be explored to identify the suitable methods and techniques for efficient classification of data. The data mining techniques are effectively used to extract meaningful relationships from these data.

2. OBJECTIVE

System predicts the oral cancer in early stages. It finds the relationship between oral cancer attributes. It will work based on the past hospital data collected. System helps doctors to know the oral cancer and relationship among attributes in early stages. It makes use of classification rules for oral cancer prediction. Implementing Apriori & genetic algorithm also improving exactness of algorithms by creating mutations of offspring. Early detection oral cancer so that it can be treated easily. Produce accurate data for further research. To build an application that is less expensive and produces accurate prediction.

3. METHODOLOGY
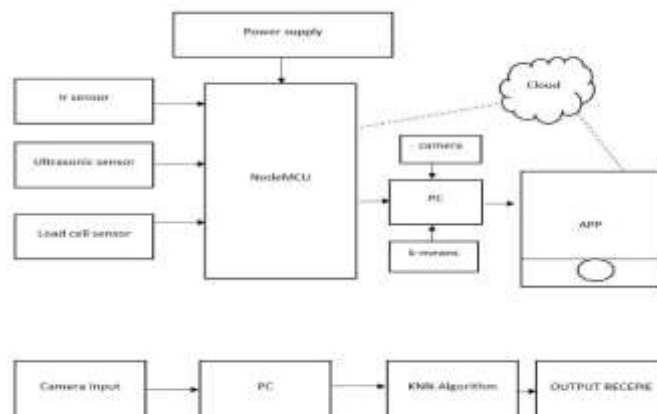
The advancement of human beings' electronics and busier lives continues to use smarter gadgets that can make life simpler. A safe and productive lifestyle is the product of smart kitchen appliances at home. One such refrigerator is the fridge. The proposed system controls the ingredients within it and notifies the user remotely with an android application of its quantity of objects. The data obtained from the sensor is analysed by the control module and sent to the cloud so that it can be accessed from the android application of the user. The smart refrigerator also offers the facility of online shopping for the scarce goods inside with quantity tracking of goods. The provision of a recipe recommendation based on the vegetables present in the basket is carried out by a simple machine learning algorithm that classifies the vegetables based on the colors, recommending a specific recipe in turn.

The user is provided with an android application which displays the quantity of items present inside the refrigerator. The food items like eggs, milk and bread are monitored using IR sensors, Ultrasonic sensor and load cell respectively. The calibrated sensors continuously read the data and compare them with threshold values. The controller classifies the sensors value as below or above the threshold and stores the same in an online cloud platform. The continuously updated data is retrieved in the mobile app designed using android studio. Based on the data displayed in the app, user is allowed to order the required items from online store with a single click. A recipe suggestion feature is provided.

A camera module is attached above the vegetable tray, present inside the refrigerator. When the user clicks on a particular button a request through MQTT client is sent to the camera module which captures an image of the vegetables. Color and number of edges from these images are extracted and K-means classifier algorithm is applied on them to identify the vegetables. K-means is a type of clustering which partitions n data into k groups in which each data belongs to the group with nearest expectation. The identified vegetables are retrieved in the app and the combination of these vegetables is used to suggest a recipe.

SOFTWARE REQUIREMENTS
- Windows 8 and above
- Python
- HTML, CSS, JS, Bootstrap
- IDLE and MS Visual Studio

PYTHON
Python is an interpreted, high-level and general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

OPENCV

OpenCV (Open-source computer vision) is a library of programming functions mainly aimed at real-time computer vision. Originally developed by Intel, it was later supported by Willow Garage then It see (which was later acquired by Intel). The library is cross-platform and free for use under the open-source BSD license. Officially launched in 1999 the OpenCV project was initially an Intel Research initiative to advance CPU-intensive applications, part of a series of projects including real-time ray tracing and 3D display walls.

NUMPY
NumPy is a Python package which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object; provide tools for integrating C, C++ etc. It is also useful in linear algebra, random number capability etc. NumPy array can also be used as an efficient multi-dimensional container for ge-

neric data. Now, let me tell you what exactly is a python numpy array. NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary datatypes can be defined. This allows NumPy to integrate with a wide variety of databases seamlessly and speedily.

PANDAS

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data. Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyse.

MATPLOTLIB

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plot into application using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on the state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.

HTML

The Hyper Text Markup Language, or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript.

Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document. HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page

CSS

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple web pages to share formatting by specifying the relevant CSS in a separate .CSS file which reduces complexity and repetition in the structural content as well as enabling the .CSS file to be cached to improve the page load speed between the pages that share the file and its formatting.

JAVA SCRIPT

JavaScript often abbreviated as JS, is a programming language thatconformstothe ECMAScript specification. JavaScript is high-level, often just-in-time compiled, and multi-paradigm.Ithas curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions. Alongside HTML and CSS, JavaScript is one of the core technologies of the World Wide Web. Over 97% of websites use it client-side for web page behavior, often incorporating third-party libraries. All major web browsers have a dedicated JavaScript engine to execute the code on the user's device.

HARDWARE REQUIREMENTS

- Processor   :   Intel i5 2.53GHz
- Hard Disk   :   30GB
- Ram   :   4 GB or above

THREE TIRE ARCHITECTURE

A 3-tier architecture is a type of software architecture which is composed of three "tiers" or "layers" of logical computing. They are often used in applications as a specific type of client-server system. 3-tier architectures provide many benefits for production and development environments by modularizing the user interface, business logic, and data storage layers. Doing so gives greater flexibility to development teams by allowing them to update a specific part of an application independently of the other parts. This added flexibility can improve overall time-to-market and

decrease development cycle times by giving development teams the ability to replace or upgrade independent tiers without affecting the other parts of the system.
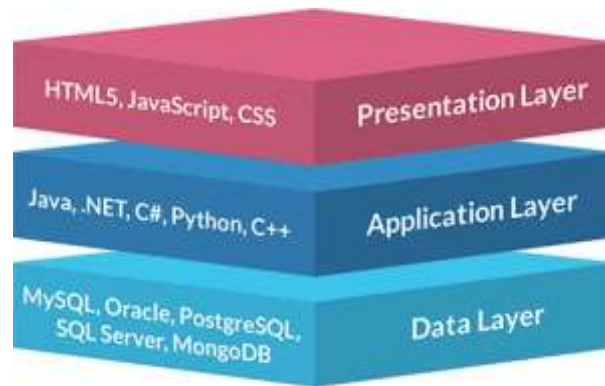
PRESENTATION LAYER

 The presentation tier is the front-end layer in the 3-tier system and consists of the user interface. This user interface is often a graphical one accessible through a web browser or web-based application and which displays content and information useful to an end user. This tier is often built on web technologies such as HTML5, JavaScript, CSS, or through other popular web development frameworks, and communicates with others layers through API calls.

MIDDLE TIRE

The Middle Tier or Business Logic layer consists of before delving further into the practical matters of application gaining a clear understanding of these concepts now will provide a sound foundation on which to build further knowledge.

DATA TIRE

The Data Tier layer consists of database such as the data tier mainly concentrated on manipulating the data using a database management system. Here we implement the data tier using MYSQL. Different varieties of data base are Oracle, SQL.

5.CONCLUSION

We can conclude that, this project provides early review of oral cancer disease. Different algorithms and techniques used for detecting oral cancer. The result of this project is to detect the cancerous cells in the oral cavity and to classify the cancer affected position to give results such as in percentage for easier approach to the doctors to start their treatment efficiently.

FUTURE WORK

In the future Innovation in diagnostic features of tumours may play a central role in development of efficient treatment methods for Oral cancer affected patients. Also, shall involve applying data mining technique to diagnose and identifying the stages, and treatments of oral cancer. In future better techniques can be incorporated with the present research work for less complexity and better learning adaptability. Moreover, better fuzzy techniques could also be used to improve the classification rate and accuracy.

- Deep learning approach for the oral cancer detection.
- Lesion based analytic model to understand its nature of spread.
- Pickle model creation for data segmentation improvisation

REFERENCES

[1] M. Sakku Vidhya, Dr. S. Shajun Nisha, Dr. M. Mohammed Sathik, https://www.ijert.org/denoising-the-ct-images-for-oropharyngeal-cancer-using-filteringtechniques, 2020

[2] Shweta Suresh Naik, Dr. Anita Dixit"Cancer Detection using Image Processing and Machine Learning" https://www.ijert.org/cancer-detection-using-image-processing-and-machine-learning, 2019

[3] Zeeba Shamim Jairajpuri, Safia Rana, Sujata Jetley" Toward early diagnosis of oral cancer: Diagnostic utilityof cytomorphological features" https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6563646/ , 2019

[4] G.Vijaya, Dr.A Suhasini "Early Detection of Lung Cancer using Data Mining Techniques", 2019

[5] Arushi Tetarbe, Tanupriya Choudhury, Teoh Teik Teo" Survey Paper on Cancer Detection Using Machine Learning", 2019.

[6] R. D. Jayasinghe, L. P. G. Sherminie, H. Amarasinghe, and M. A. Sitheeque, ''Level of awareness of oral cancer and oral potentially malignant disorders among medical and dental undergraduates,'' Ceylon Med. J., vol. 61, no. 2, p. 77, Jun. 2016

[7] O. Kujan, A.-M. Glenny, R. Oliver, N. Thakker, and P. Sloan, ''Screening programmes for the early detection and prevention of oral cancer,'' Austral. Dental J., vol. 54, no. 2, pp. 170– 172, Jun. 2009.

[8] N. Haron, R. B. Zain, W. M. Nabillah, A. Saleh, T. G. Kallarakkal, A. Ramanathan, S. H. M. Sinon, I. A. Razak, and S. C. Cheong, ''Mobile phone imaging in low resource settings for early detection of oral cancer and concordance with clinical oral examination,'' Telemed. eHealth, vol. 23, no. 3, pp. 192–199, Mar. 2017.