

# Healthcare Data Fusion

Adithya Jayaprakash Pillai  
 Computer Science and Engineering  
 Iahia College of Engineering and Technology  
 Muvattupuzha, Kerala, India

Nirmal T R  
 Computer Science and Engineering  
 Iahia College of Engineering and Technology  
 Muvattupuzha, Kerala, India

Asif Ali  
 Computer Science and Engineering  
 Iahia College of Engineering and Technology  
 Muvattupuzha, Kerala, India

Asni K K  
 Computer Science and Engineering  
 Iahia College of Engineering and Technology  
 Muvattupuzha, Kerala, India

**Abstract**—The system proposed in this paper is a smart healthcare solution for monitoring and bringing about insightful data from diverse data types that are produced by a patient using the concept of data fusion by implementing various machine learning techniques. As a growth is seen in the increasing number of clinical devices and the swift advancement in the field of healthcare such a solution helps one to develop a more accurate and appropriate understanding of a particular disease or an outbreak.

can be tracked and controlled and also would improve the quality of life .

## I. INTRODUCTION

Today in the modern world where healthcare applications and devices are being used to scale up the quality of the kind of healthcare service provided, large amounts of data are being generated at unmatched speed and great volume. This data can be used for extracting statistically advanced insights. The use of body sensors are required in healthcare applications and devices to transmit data based on a patient’s healthcare level and network. This is where data fusion and machine learning comes into play. The primary aim of such a solution is to reduce unwanted time and resources being spent in the wrong direction while we can get faster and more accurate results to take clinical decisions in the right direction.

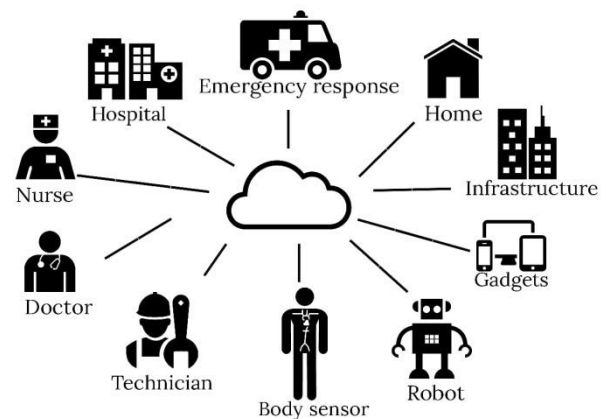


Fig. 1. Hospital Network

## II. PROPOSED SYSTEM

The increasing speed at which large amount of medical data collected across various platforms has put a rigorous challenge to healthcare industry for careful integration and implementation of this data on a whole new level. Therefore, this paper suggests that revolution in healthcare is further needed by grouping together genetics and health informatics to promote personalized care and more efficient and accurate treatments. Here we propose a system which uses NLTK, Deep Learning and Machine Learning to process and analyse data which are produced by different elements in the healthcare industry such as Doctor, BSN’s, Nurse etc. These data are collected and then processed using a series of steps like Regrouping, Recombination, Estimation, Aggregation etc finally generating a report which shows the present health condition of a patient from time to time helping Clinical officials to make a better decision on the kind of treatment and care that is to be provided to the patient. This can be applied on different levels so that different disease outbreaks

Using Neural Networks we can sort and process data in a more precise way which brings out the best results. Neural Networks produce results after multiple computations by reducing the error rate. This helps in getting the most accurate and best results. Data such as Text data and Medical images are fused accordingly to get the most optimal decision support

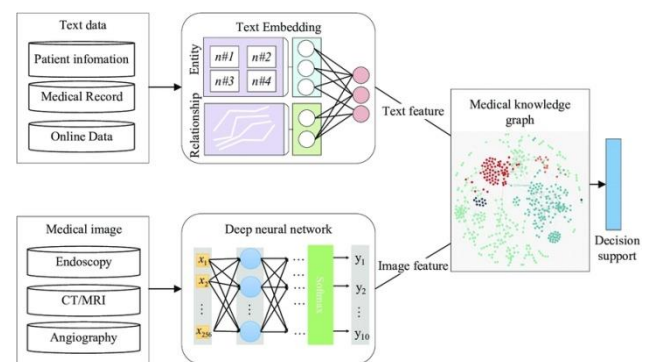


Fig. 2. Learning process of the system

### III. SYSTEM ARCHITECTURE

#### A. Data Regrouping

Data regrouping is the process of grouping two or more data sets into a single data set. Usually, such a process is must to do when there is raw data that is stored in multiple files, worksheets, or data tables, which of all is to be analyzed.

#### B. Data Recombination

Finding similar images and removing the duplicated images, similar images mean images with the same content. Removing data having the same hash value, hash value created with MD5 algorithm. Correcting the merged data with similar data and storing the weightage of the data.

#### C. Remove Redundant Data

Checking and removing errors from data. Removing noise from the images and finding the edges with edge detection algorithms. Removal of data and columns having None values.

#### D. Data Aggregation

Data aggregation is the process in which raw data is searched, gathered, and presented in a summarized, report-based form. The data may be gathered from multiple data sources with the intent of combining these data sources into a summary for data estimation.

#### E. Data Estimation

Estimation is a division of statistics and signal processing that determines the value of parameters through measured and observed empirical data. The process of estimation is carried out in order to measure and diagnose the true value of a function or a particular set of proportions. Data estimation might be viewed as a set combination wherein the larger set is retained.

#### F. Data Fusion

Data Fusion processes are often categorized as low, intermediate, or high, depending on the several sources of raw data to produce new raw data. The expectation is that the fused data is more informative and synthetic than the original inputs. Multiple fusions are carried out using boosting, voting and bagging techniques.

#### G. Data Compression

The data that is produced after fusion is considered to be large in size. So data compression techniques are applied on the fused data to make it easier for transferring to different storage and also for the production of reports.

#### H. Report Formatting

A particular pattern is found out to produce a rough report of the data. Also the dictionary which contains the weightage of the different words is taken into consideration while creating the same.

#### I. Report Production

Producing the report based up on the data. Based upon the report formatting scheme(analyzed data) a new and more structured summarized report is created.

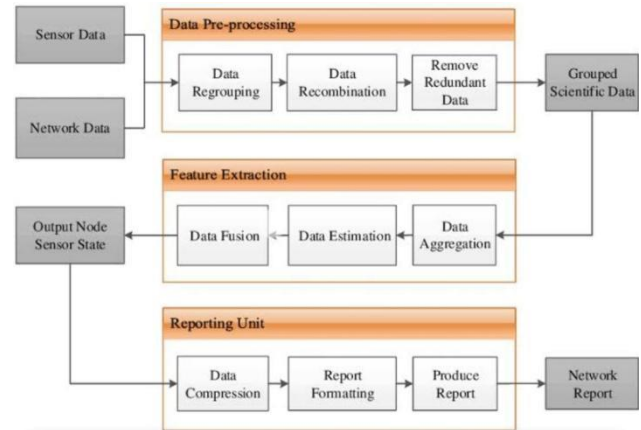


Fig. 3. System Architecture

### IV. PRELIMINARIES

The different techniques and algorithms that are used during data fusion are:

#### A. Deep Learning

Deep Learning is a part of machine learning which works on the concept of artificial neural networks. It's more alike the functionality of the human brain in processing information and also in structure.

#### B. Sobel Edge Detection

The Sobel edge detection algorithm is an algorithm that is uses Sober filter for edge detection in images and graphics. It basically work by calculating the gradient of image changes at each pixel within the image. It calculates the direction of the largest increase from light to dark and then the rate of change in that direction. Finally, it depicts how smoothly the image changes at each pixel, and therefore how the pixel represents an edge.

#### C. Machine Learning

Machine learning is one of the best use cases of artificial intelligence (AI) that enables systems to learn automatically and improve from their experiences without being told what to do when an event occurs. There are different types of machine learning algorithms that can be used for different purposes according to the need.

#### D. NLTK

The Natural Language Toolkit, is a package for understanding Natural Language Processing for the English language which is primarily written in Python developed by Steven Bird and Edward Loper. It basically helps in language processing and understanding linguistics and artificial intelligence.

### V. LITERATURE SURVEY

Based on the survey conducted, we found out these papers of our best interests :

According to Muhammad Muzammala, Romana Talata, Ali Hassan Sodhrob and Sandeep Pirbhulalcin "A multi-sensor data fusion enabled ensemble approach for medical data" [1] , Wireless Body Sensor Network (BSNs) are wearable sensors with varying sensing, storage, computation, and transmission capabilities. When data is obtained from

multiple devices, multi-sensor fusion is desirable to transform potentially erroneous sensor data into high quality fused data.

According to ShridaKalamkar and Geetha Mary A, in "Clinical Data Fusion and Machine Learning Techniques in Healthcare" [2], In clinical decisions, generally, one source of information may not be able to achieve an accurate decision. For accurate analysis, heterogeneous data can be fused, which in turn helps to develop a more understanding of a particular disease. Fusing data from diverse sources like clinical repositories, sensory devices, historical or textual data is vital for both patients and healthcare providers.

According to RustemDautov, Salvatore Distefano and RajkumaarBuyyya in "Hierarchical Data Fusion for Smart Healthcare" [3], proposes a distributed hierarchical data fusion architecture, in which different data sources are combined at each level of the IoT taxonomy to produce timely and accurate results. This way, mission-critical decisions, as demonstrated by the presented Smart Healthcare scenario, are taken with minimum time delay, as soon as necessary information is generated and collected.

According to Qingguo Zhang, Bizhen Lia, Ping Cao, Yong Sang, Wanli Huang and Lianyong Qi in "Multi-Source Medical Data Integration and Mining for Healthcare Services" [4], novel multi-source medical data integration and mining solution for better healthcare services, named PDFM (Privacy-free Data Fusion and Mining). Through PDFM, we can search for similar medical records in a time-efficient and privacy-preserving manner, so as to offer patients with better medical and health services.

## VI. CONCLUSION

The growth of every field in today's world gives an importance to advanced healthcare. This system makes an improvement in the time needed to provide an effective treatment. Providing the best facilities at the right time is the most important when it comes to the medical field. There are other existing systems for diagnoses but the proposed system helps in more personalized healthcare.

## REFERENCES

- [1] Muhammad Muzammala, Romana Talata, Ali Hassan Sodhrob and Sandeep Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data"
- [2] ShridaKalamkar and Geetha Mary A, "Clinical Data Fusion and Machine Learning Techniques in Healthcare"
- [3] RustemDautov, Salvatore Distefano and RajkumaarBuyyya, "Hierarchical Data Fusion for Smart Healthcare"
- [4] Qingguo Zhang, Bizhen Lia, Ping Cao, Yong Sang, Wanli Huang and Lianyong Qi, "Multi-Source Medical Data Integration and Mining for Healthcare Services"
- [5] JunhaiZhai, Sufang Zhang and Chenxi Wang, "The Classification of Imbalanced Large Data Sets based on MapReduce and Ensemble of ELM Classifiers", Springer-Verlag Berlin Heidelberg, Springer 2015
- [6] NongyaoNai-arun and PunneeSittidech, "Ensemble Learning Model for Diabetes Classification", Faculty of Science, Naresuan University, Phitsanulok, Thailand, Advanced Materials Research Vols. 931-932 pp. 1427-1431, Trans Tech Publications, Switzerland, 2014
- [7] Ping Deng, Honghun Wang, Shi-Jinn Horng, Dexian Wang, Ji Zhang and Hengxue Zhou, "Softmax Regression by Using Unsupervised Ensemble Learning", 2018 9th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), IEEE 2018
- [8] Prashant Dhotre, SayaliShimpi, Pooja Suryawanshi, Maya Sanghati, "Health Care Analysis Using Hadoop", Department of Computer Engineering, SITS, Narhe, IJSTR 2015
- [9] Priyanka Dhaka, Rahul Johari, "HCAB: HealthCare Analysis and Data Archival using Big Data Tool", Indraprastha University, New Delhi, India, IEEE 2016
- [10] Shikha Mehta, Priyanka Rana, Shivam Singh, Ankita Aharma, Parul Agarwal, "Ensemble Learning Approach for Enhanced Stock Prediction", Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India, IEEE 2019
- [11] Shuaichao Gao, Jianhua Dai, Hong Shi, "Discernibility Matrix-Based Ensemble Learning", 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, IEEE 2018
- [12] Tanmoy Chakraborty, "EC3: Combining Clustering and Classification for Ensemble Learning", Dept of CSE, IIIT Delhi, India, 2017 IEEE International Conference on Data Mining, IEEE 2017