A Smart Multimodal Biomedical Diagnosis Based on Patient's Medical Questions and Symptoms

Dr Vijaya Gunturu[1], R.Krishnamoorthy[2], M. Amina Begum[3], Dr R Jayakarthik[4], Kazuaki Tanaka[5], Janjhyam Venkata Naga Ramesh[6]
Department of ECE, School of Engineering, SR University, Warangal, Telangana[1]
Centre for Advanced Wireless Integrated Technology, Chennai Institute of Technology, Chennai, India[2]
Department of ECE, Oxford College of Engineering, Venmani, Thiruvannamalai District, India[3]
Department of Computer Science, Saveetha College of Liberal Arts and Sciences SIMATS, India[4]
Kyushu Institute of Technology, Japan[5]
Dept.of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522502,India[6]
[1]vijaya.gunturu@sru.edu.in, [2]krishnamoorthyr@citchennai.net [3]amina.gtec@gmail.com, [4]drrjayakarthik@gmail.com, [5]kazuaki@ics.kyutech.ac.jp [6]jvnramesh@gmail.com

Abstract: The exponential increase of health-related digital data has given machine learning algorithms a newfound ability to generate more meaningful insights. Information such as diagnosis, treatments, and prescriptions are all part of digital health data. In order to better care for their patients, healthcare providers provide crucial diagnostic services. Mistakes in diagnosis, however, lead to the patient receiving harmful treatment too soon or too late. In order to reduce the likelihood of clinical cognitive errors, computer-aided diagnosis techniques have been developed. The proposed approach makes utilize of a massive health-related dataset, which comprises many unstructured patient questions written in various Arabic dialects, as well as symptoms reported by general-practitioners (GPs). System components include a combination of machine learning models that have been trained using either patient symptoms or patient medical inquiries. Machine learning (ML) strategies, and variations of the Multilayer-Perceptron (MLP) classifier have all been utilized in trials as feature representation techniques and machine learning classifiers. We also discuss the technical and analytical hurdles, as well as the most important new applications, that this research opens up. Possibilities in areas such as digital clinical trials, telehealth, pandemic surveillance, digital twins, and virtual health aides are discussed. We also provide an overview of the data, modeling, and privacy obstacles that must be surmounted before the health care industry can fully benefit from multimodal AI. With a classification accuracy of 94.9%, the combined results of the two modalities demonstrate promising prediction potential. The results show promise for using the algorithm to predict possible diagnoses of patient illnesses depends on the given symptoms and queries, which can help doctors, make more informed judgments.

Key Terms: Multimodal-diagnosis, Machine-learning, Deep-learning, Digital health, Computer-aided diagnosis, biomedical-diagnosis

I. INTRODUCTION

In the recent decade, developments in portable and wearable sensor technologies, IoT-based medical tools, and big data analytics via artificial intelligence and machine learning have fueled a meteoric rise in the area of smart health [1]. Bigdata service platforms, which offer the infrastructure for large-scale data space and processing, have helped to speed up the industry as well [2]. Constant multimodal monitoring of physiological conditions is just one aspect of the patient-generated health-data(PGHD) that has been made possible by the widespread availability and use of wearable-wireless sensor tools. Healthcare can be shifted toward prevention and early disease diagnosis with the help of digital health because of the big data analytical technologies used, especially ML &DL [3].

While smart-health technologies have made it easier to keep tabs on a wide range of physiological states, one issue that has arisen is how to ensure that all of the devices involved in monitoring a patient's health are in sync with one another. The accuracy of the timing of physiological signals is crucial to their subsequent applications [4]. To begin with, it is the backbone of comparing new technology to industry standards. The digital health toolkit is always being updated with new features and data modalities. Comparisons with data obtained at the same time from gold-standard medical devices are used in validation studies to verify their accuracy [5]. The core of the validation is the ability to precisely match the timestamps of various signals. Second, an increasing number of research use supplementary data from multimodal physiological signals to uncover novel insights and enhance system performance across a range of tasks. Again, precise time alignment is the bedrock upon which such synchronous multimodal datasets are constructed [6].

It is tricky to determine the likeness metric for the time-matching techniques when synchronizing multimodal signals since their morphology may differ. In addition, there is no universal truth available to verify the efficacy of the algorithm while dealing with signals from various devices [7]. In an effort to develop a time alliance key for multimodal physiological signals, numerous investigations have been conducted. To identify the best temporal alignment between two signals, they utilized ML models. When two time-series are calculated with diverse sample rates, the DTW algorithm is known to experience the singularity problem. By exploiting local knowledge in the signal from the prescribed actions, the authors ease this problem and show superior performance compared to two previous classical DTW algorithms [8]. Another unique method developed by Liu et al., dubbed Phyio2Video, turns the signal alignment problem into a video frame configuration task by extracting spectral features from various physiological data. Using a DCNN to achieve nonlinear encoding of the feature films, the authors were able to handle the video frame alignment task and utilize a canonical correlation loss to calculate the final alliance of two signals [9].

There are a number of different prediction-related learning methods that can be loosely categorized as model-based or data-based. The former rely on preconceived notions of how the data should evolve, while the latter make use of the plethora of digital resources at their disposal and the ever-improving performance of AI strategies to train from the data itself [10]. From an artificial intelligence standpoint, the three basic fusion strategies that combine information from several modalities are early fusion, joint fusion, and late fusion. The first method combines the features from each modality into a single feature vector before presenting it to the learner; the next method combines the modalities at the hidden and entrenched levels; and the third method aggregates the predictions from the various modalities [11].

To aid in the clinical diagnosis and evaluation of the efficacy of a wide range of major diseases, DL offers exact strategies for processing massive data. Since medical image analysis is a crucial part of current medical imaging technology [12], it is imperative that this long-standing scientific challenge be resolved as soon as possible. The utilize of DL for multi-modal medicinal image fusion has the potential to greatly enhance the diagnostic and evaluative utility of medical images. Videos, for instance, can be broken down into its component parts, which include static text, static images, and static speech [13]. Evidence from the field suggests that multi-model approaches to processing data outperform their single-model counterparts [14]. To enhance image quality while keeping certain details intact, medical professionals often resort to multi-modal medical image fusion [15]. Because of the many fascinating disciplines involved, medical image fusion has found widespread use in clinical practice. These disciplines include image-processing, computer-vision, pattern-recognition, ML, and AI. Medicinal image fusion provides doctors with new perspectives on lesions.

Most current uses of AI in healthcare have focused on solving specific problems with a single data-modality, such as a computed-tomography (CT) scan or an image of the retina. When making a diagnosis, prognosis assessment, or treatment plan, however, professionals interpret data from a wide variety of sources and modalities [16]. Furthermore, present AI assessments are often one-off snapshots, depending on a moment in time when the appraisal is performed, and therefore not seeing health as an unremitting condition. However, AI models should be capable to exploit any and all datasources, including ones that are occupied to most doctors [17]. We analyze the potential benefits of such multimodal datasets in healthcare, along with the major obstacles they provide, and some possible approaches to overcoming them in this review. The fundamentals of artificial intelligence and machine learning will not be covered here, but they are extensively covered elsewhere.

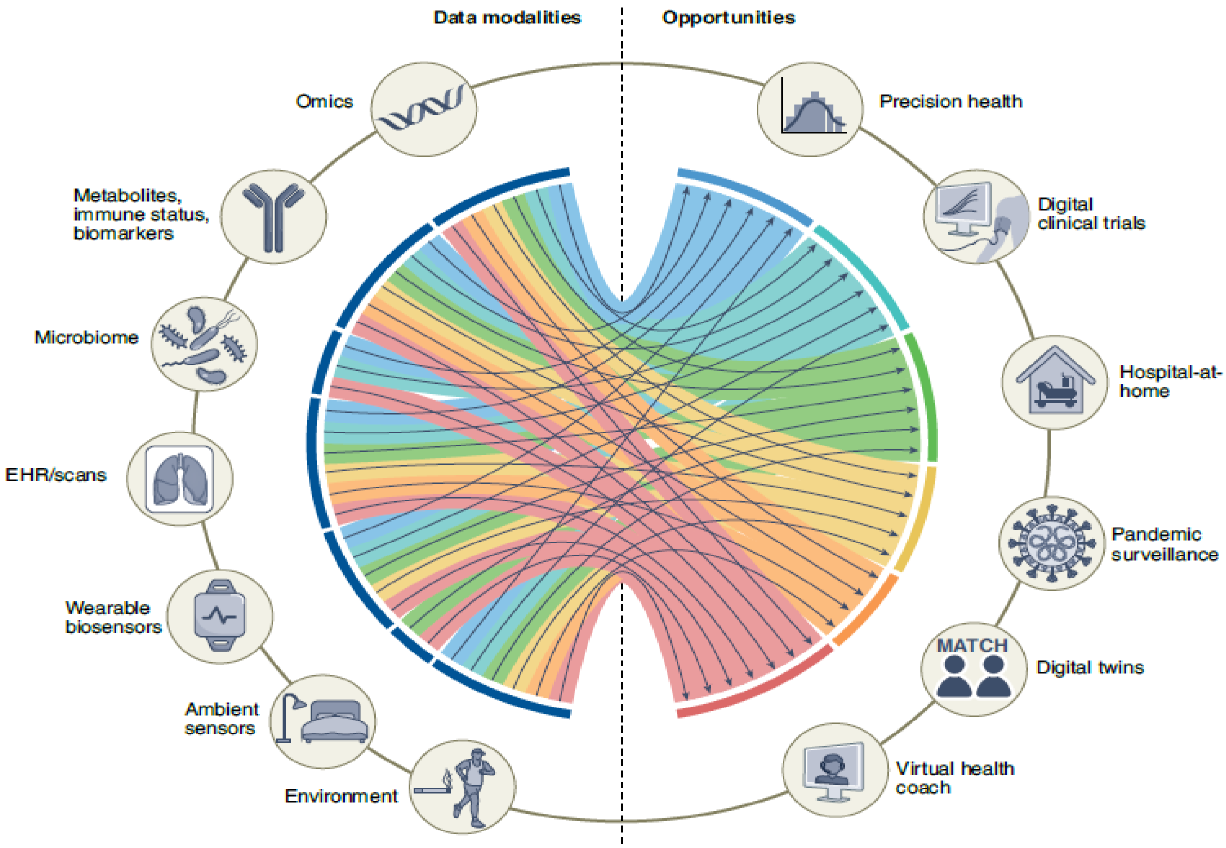

Fig. 1. Data modalities and opportunities for multimodal biomedical AI

The issue and its impetus are depicted in Figure 2. In this chapter, we use a multimodal classification strategy based on machine learning techniques to the issue of detecting potential diagnoses. This paradigm is projected to have a number of benefits, including, first, facilitating accurate identification at an early stage, which is difficult because symptoms at this time are often vague or inconsistent [18]. Next, the capacity to assimilate vital information such as a patient's medicinal history or an allergy, the absence of which complicate the diagnostic procedure and sometimes leads to an incorrect illness differentiation [19]. Third, it helps with the time-consuming and error-prone process of mapping clinical notes to diagnoses according to the International Classification of Diseases (ICD) [20]. The suggested classification scheme integrates several different approaches. As a result, it integrates data, features, scores, and decisions from different sources, each of which contributes something unique. The effectiveness of the learning algorithm can be enhanced by combining data from several modalities. Integrating data from a person's facial signs, speech, behavior, and physiological signals can help a machine learning model better understand that person's emotions [21]. Patients' inquiries and doctors' descriptions of symptoms form the basis for the proposed multimodal machine learning system. In this setup, two ML strategies are constructed separately for every modality, and then their output is integrated to form a whole.

Text vectorization methods, which convert words in a document into numerical values, process the patients' inquiries. The TF-IDF and the hashing-vectorizer are two such methods that focus mostly on syntactical aspects of a document. Also, the embedding models (like Doc2vec embedding) that can be used to glean the documents' hidden meanings [22]. On the other hand, doctors' notations of ICD-10 codes for patients' symptoms constitute organized data. The One-Versus-Rest (OVR) method is used to map consultations to their accurate diagnosis, which is formulated as a multi-class classification. In order to solve multi-class classification issues, OVR employs a heuristic method that generalizes the capabilities of binary-based machine learning techniques. Experiments have been conducted using several machine learning classifiers and comparing them separately according to each modality [23]. The LR, RF, SGDClassifier, and MLP classifiers were utilized, and their details are provided later in the paper. Various methods, including ranking, summing, and multiplication, are used to merge the two models' final results. The suggested model is tested for precision, efficiency in inference and loading, and manageability in terms of classification model size. The designed diagnosis approach accomplished an impressively high rate of accuracy (84.9%) in its categorization results.

The suggested method's primary contributions are:

• Combining organized clinical data with unstructured free-text consultations to create a diagnosis decision support system.

• The formidable difficulty of designing a system to accommodate the wide variety of Arabic dialects spoken today. To further aid physicians in making sound diagnoses, the suggested technology will be integrated into the digital health platform.

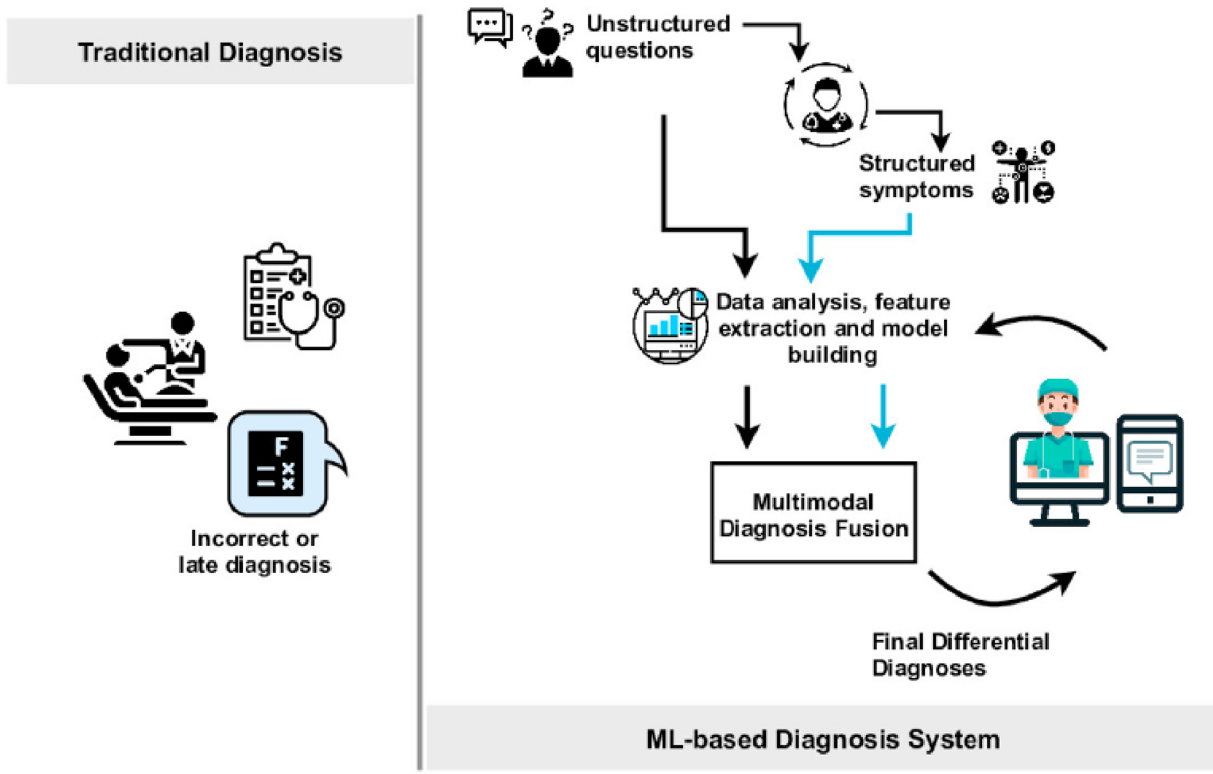

Fig.2. The conventional and ML-based discrepancy diagnosis structure

The rest parts of the chapter are structured as pursues. Section2 presents recent relevant work in DL &ML-based differential diagnosis systems. In Section 3, we discuss the technique, which includes the data gathering, preprocessing, features-extraction, proposed QSDM architecture, and assessment criteria. Section 4 offered the investigational conditions, the actual trials, and an argument of the outcomes. Section 5 concludes with the results and recommendations for further research.

II. RELATED WORKS

A thing's way of existence, experience, or expression is its modality. A multimodal research challenge is one that involves more than one possible solution. At the same time, modes can be defined in a wide variety of ways. Data collected in two languages or fewer than two different conditions, for instance, can be considered two modes [24]. One of the keys to understanding the world around us is picking up on several signals at once. Images typically have accompanying labels and explanatory text, and articles frequently include visuals to help convey the article's main point. Since each mode has its own set of statistical characteristics, MMDLM can be utilized to interpret and comprehend multi-source modal information through the lens of deep learning. Figure 3 shows the variety of existing models [25].

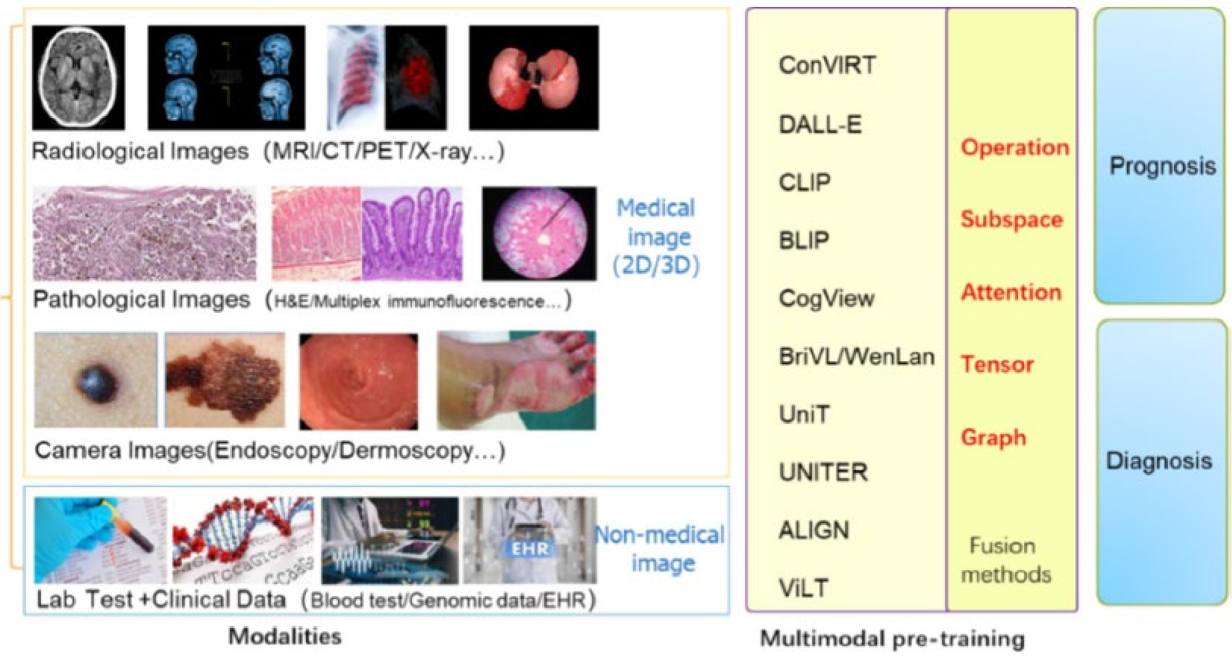

Fig. 3. Structure of the related works

At the moment, researchers are digging deep into multi-modal learning for pictures, videos, sounds, and words [26]. For the purpose of learning multi-layer displayers and the pensiveness of data prior to its conversion into high-level intangible properties of the system, a lot of focus is currently on deep neural networks. Classification, segmentation, detection, and localization are just few of the medical applications that have benefited greatly from research into image analysis [27]. Segmentation, anomaly-detection, illness categorization, computer-aided analysis, and picture rescue are just few of the medical image processing tasks that have seen widespread application of deep convolutional networks. Clinical applications have relied on medical imaging as a diagnostic tool for quite some time. The field of medicinal imaging has profited greatly from modern expansions in device design, security software, computer power, and data-storage capacity [28]. Findings demonstrated that the interpretable AI-aided analysis greatly enhanced the diagnostic accuracy of doctors, enhanced the irrefutable use of its supplementary verdict, and generated novel hypotheses for future studies of clinical translation. As medical IT and tools have advanced, so too has the volume and variety of medical data that has been available [29]. Based on its content and structure, medical data can be divided into three primary groups:

1. Clinical text data is mostly structured test data like hemoglobin and urine routine and unstructured text data like patient complaints and clinician-recorded pathologic texts.

2. Imaging data (ultrasound, CT, MRI) and waveform data (ECG, EEG) are examples of images and waveforms.

3. Biomics data, which can be further categorized at the molecular level into genomic, transcriptomic, proteomic, and other types.

In [30], authors used MIMIC-III to create a NN-based approach for clinical note diagnosis. The top 10 classification was 80% more accurate than the top 50. Ref. [31] automated clinical document ICD-10 mapping. BOW, TF-IDF, Word2Vec, LSTM, and CNN were combined into the SVM algorithm. Deep learning classifier outperformed. In [32], they presented an automated pathology report diagnosis classification. TF-IDF extracted features for linear-SVM, XGBoost, and LR. XGBoost had the highest f1-score (92%). In [33], authors used emergency department data to automate mental state detection. SVM, NB, RF, CNN were compared. DL performed best with 98.1% accuracy. In [34], authors created a web-based ML system for mental disease analysis. The tool matches symptoms to an ICD-10 disorder. The TF-IDF feature vectorizer was used to train the KNN classifier. To diagnose Alzheimer's and Vascular Dementia using machine learning. Adaptive neuro-fuzzy inference has 84% accuracy.

In [35], researchers created a ML strategy to diagnose and predict major depressive and bipolar disorder. The model scored 97% area under the curve for a limited dataset. They constructed a ML approach to predict schizophrenia-bipolar disorder. It incorporates multi-domain immune indicators from 513 diseases. Based on 16,114 instances, authors [36] developed a deep CNN for disparity skin disease investigation. It predicted 419 skin disorders and recognized 26. It has 73% top-one precision, whereas three professional dermatologists had 63%. In [37], creators applied DL to ultrasound images to differentially diagnose COVID-19. The model classified COVID-19, pneumonia, and healthy pictures with almost 90% accuracy. The previous investigations showed possible efforts to apply discrepancy analysis to improve physician decision-making. They showed that Arabic has no such mechanisms. MENA clinical diagnosis decision support systems need more research [38].

Multimodal DL uses deep-learning to combine many data kinds. Imaging aids medical diagnosis. Since clinical diagnosis requires processing significant volumes of data, single-mode medical images provide limited information. Deep learning-based multimodal medical image fusion may extract and combine feature information from multiple modalities, improving medical image diagnostics and evaluation. It has gained popularity due to its benefits.

This study addressed these literature gaps:

1. Insufficient medical imaging multi-modal data fusion literature to organize and summarize studies.

2. Lack of clinical expertise and data scientists and data analysis algorithm knowledge.

3. Lack of multimodal DL bibliometric analysis in medicinal imaging.

III. PROPOSED METHODOLOGY

The methodology includes data collection and preprocessing, feature extraction for questions, classification model creation, and model evaluation. Figure 4 summarizes the process.

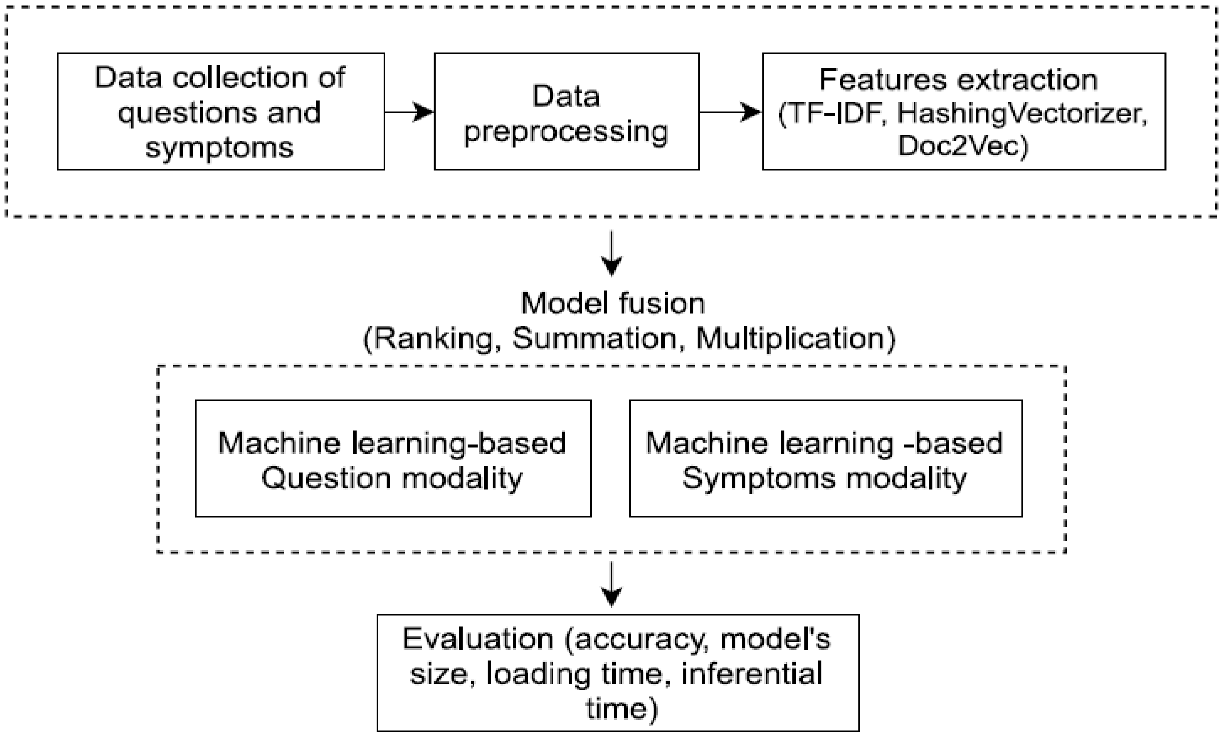

Fig.4. Proposed approach concept diagram

3.1. Data-collection &preprocessing

The total amount of information obtained through Altibbi consists of 263,867 questions (consultations), each of which is escorted by symptoms and queries. The overall quantity of diagnoses is 8,410, whereas the number of symptoms comes in at 7,324. Each appointment is accompanied by a number of symptoms and a number of diagnoses, despite the fact that some of these conditions only appear seldom. In the first place, we got rid of the diagnoses that came up less than 20 times throughout all of the consultations [39]. After that, the consultations that had resulted in a lack of a diagnosis were eliminated. As a result, the total amount of diagnoses was 2368, while the quantity of consultations came to a total of 246,814. A comparison of the number of diagnoses with the total number of consultations is presented in Figure 5. It is quite obvious that the majority of consultations concern a single diagnosis. During this time, a number of preprocessing activities are carried out in order to spotless and organize the data in preparation for the predictive approach [40].

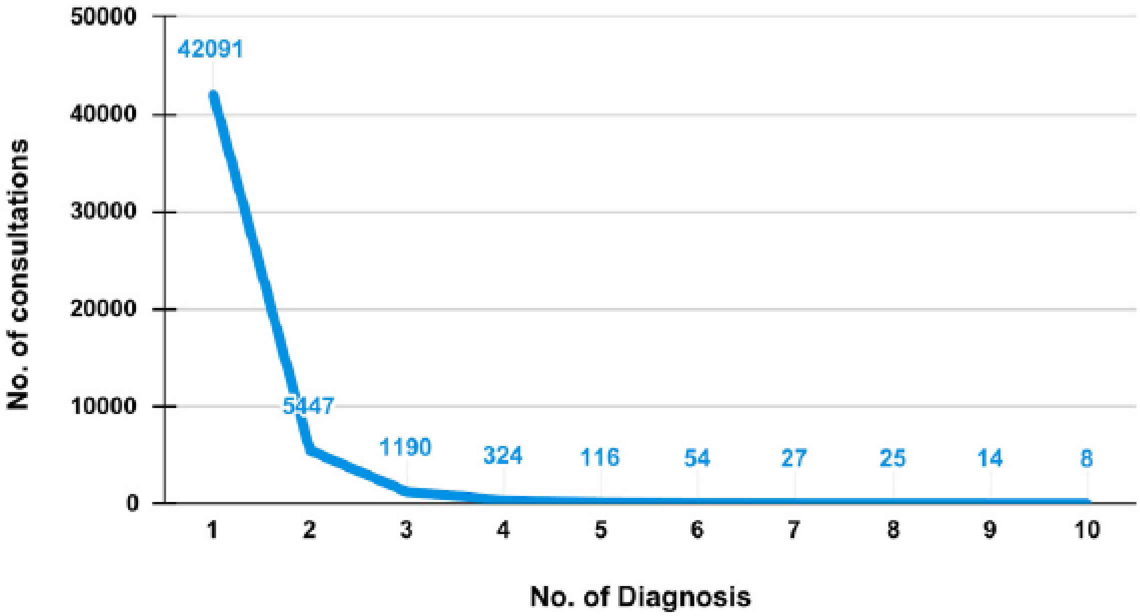

Fig. 5. The association amid the quantity of consultations and investigation

3.2. Feature extraction

Vectorization is the primary method utilized in the process of extracting features from the textual material. The practice of translating textual records into arithmetical feature-vectors is referred to as "vectorization." Several methods, such as TF-IDF, the hashing-vectorizer, and word-embeddings, have been suggested in research that has been published so far [41].

3.3. Question-symptom-diagnosis model (QSDM)

This primary focus is on outlining the process that went into designing the QSDM strategy. It is a combination of two diagnostic approaches: initially, examines the symptoms and organizes them into four distinct diagnoses based on their characteristics. Because offering more than 4potential scrutiny is likely to be confusing to the patient, we have decided to limit the number of diagnoses that can be suggested to only four. The second modality is called the question categorization modality, and it forecasts a maximum of four possible diagnoses. The final forecast is determined by integrating the results of the first and second modalities.

3.4 System architecture

As can be seen in figure 6, the QSDM is essentially the result of the combination of two distinct models: the symptom-finding and the questions model. The output of the questions design are intended to be improved by merging the results of the symptoms model with the questions model in order to aggregate informative elements from the symptoms design. It takes into account all symptoms as binary features; hence, it takes into account the complete collection of distinctive symptoms gleaned from the queries (7,324 features in total). The suite of labels (2368), each of which is showed by a binary value, is what constitutes the individual diagnoses. Eighty percent of the symptom data will be used for training, and twenty percent will be used for testing. The information is then processed by a number of different ML strategies. The learning models are constructed with the help of the training set, while the testing set is utilized in order to evaluate how well the models have performed. When it comes to dealing with multi-class categorization, the developed models use the OVR approach as their foundation. Each every model undergoes training and testing on an individual basis. However, the final anticipated diagnoses are derived from the component of this submodel that has the highest classification accuracy. The TF-IDF algorithm, the hashing-vectorizer, and document embedding were the feature extraction methods that were used individually for the questions model. The document embedding was accomplished with the help of the Doc2Vec program. The three datasets that were generated are then split into training datasets consisting of 80% and testing datasets consisting of 20%, respectively. During this time, they are being sent through OVR and into the four classifiers. Next, the results of the classifier that had the best overall performance are chosen to serve as the ultimate predictions of the question model.

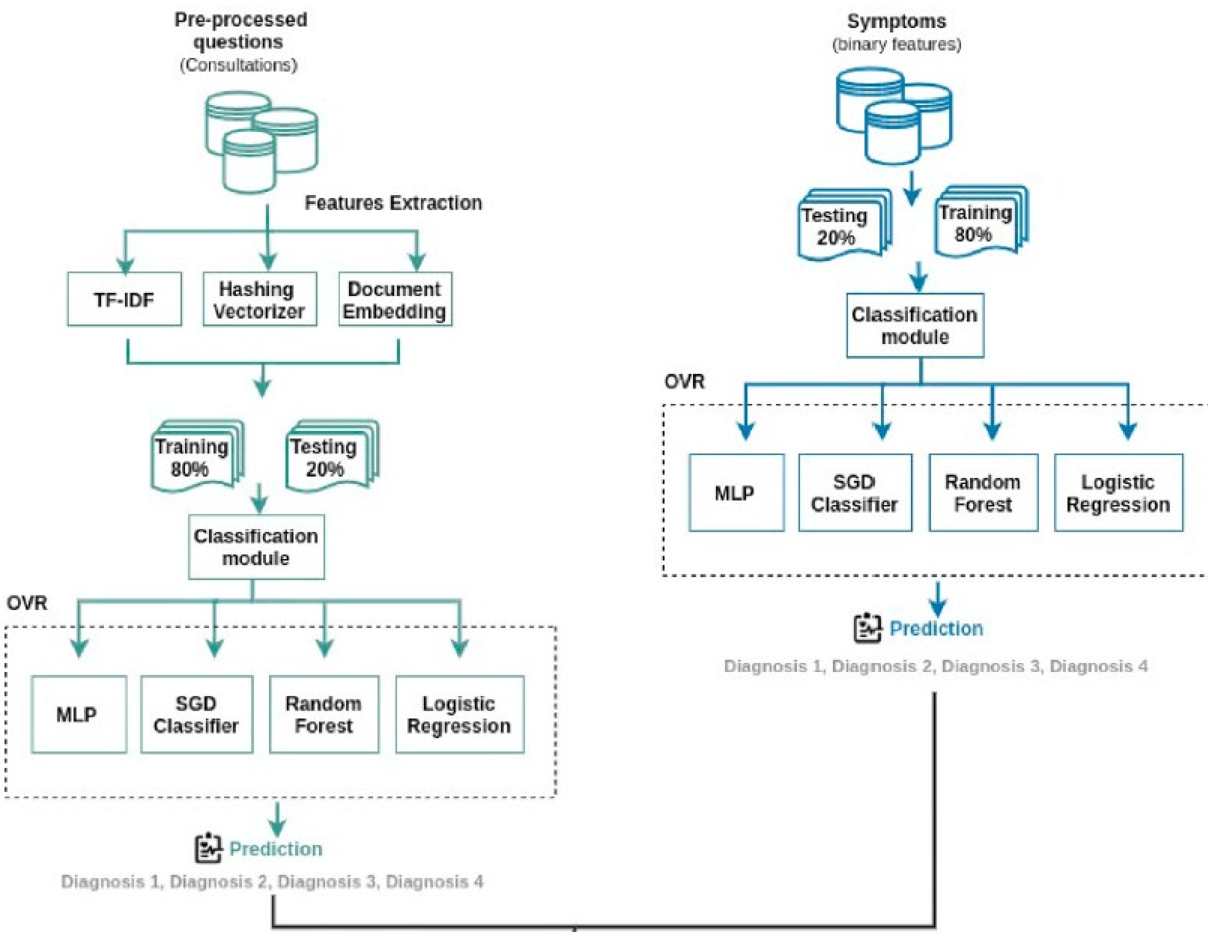

Fig. 6. Architecture of the QSDM model

3.4. Evaluation criteria

The accuracy at varying degrees of precision, the size of the technique, the loading and inferential time were the four quantitative assessment variables are examined while measuring the performance of the QSDM model. According to Equations (1) and (2), accuracy is the proportion of correct diagnoses relative to the number of diagnoses (m). Probabilities of all possible diagnosis are represented by Px in Eq. (1), where X = [x1, x2,..xn], where n is the total number of possible diagnoses. Consultation actual diagnosis (y) is substituted for the expected diagnostic (x) in Equation (2). Where (m) is the total number of potential diagnoses, in this case 4. Probabilities of all diagnoses, denoted by P, and indexes of diagnoses, denoted by j.

$\text{Argmax } P_x = \{x | \text{ if } v>z\}$ Eq. (1)

Accuracy (Acc) = $\frac{1}{m}\sum_{i}^{m}\{\ f(y) = 1|y = argmax(P_a\ )\ \}$ Eq. (2)

The precision (P) is how the accuracy is demonstrated to the audience. For instance, the accuracy at precision one indicates how well the algorithm is able to get at least one true diagnosis out of the truth diagnoses that are available. This is what is known as P_1, if you were wondering. P_2 refers to the ability of the model to identify at least two right diagnoses, whilst P_3 represents the ability to identify at least three diagnoses. The size of the model is a significant metric, particularly in light of the fact that increasing the size of the model would invariably lead to an improvement in the model's overall performance. Nevertheless, it is essential since, in circumstances in which there is a limited amount of infrastructure, it may reduce the effectiveness. Additionally, the loading time and the inferential time are two relevant measures that indicate the effectiveness of the model in delivering real-time predictions. The amount of time required to load the model onto the web is referred to as the loading time, while the amount of time required making a prediction is referred to as the inferential time.

IV. RESULTS AND DISCUSSIONS

With regard to the questions-module, this part presents a contrast of the classifiers at various feature-extraction approaches. These techniques include the TF-IDF, hashing-vectorizer, and the record embeddings. It is abundantly obvious from the data presented that all strategies produced superior outcomes when they properly anticipated a minimum of one diagnosis (P_1). According to the results presented in the table, the LR approach was the top performing classifier overall (53.7%). Even while the MLP (10) showed a minor decrease in exactness as compared with LR, it still managed to attain a very respectable 45.2% accuracy. Both the MLP (20) and the MLP (30), however, were able to attain pretty respectable results (44.0% and 41.4%, respectively). On the other hand, the SGDClassifier had the worst performance (33.5%). In terms of the scenario to forecast minimum of 2 correct diagnose (P_2), the LR fared the finest with a percentage of 40.4%, followed by the MLP (10) and the MLP (20) with percentages of 36.7% and 38%, correspondingly. In the same vein, the LR obtained the highest accuracy (39%) when it came to predicting at least three right diagnoses (P_3), followed by MLP (10), which had an accuracy of 37.9%, and MLP(20), which had an exactness of 39%, correspondingly. When constructing a machine learning model, some of the most significant factors to take into consideration are the model's size, the amount of time necessary to install it on the web, and the amount of time required to execute an inferential anticipation. Incidentally, the MLP(10) had the smallest size, at 5.2MB, in terms of the M.S., while e RF had the largest size, at 17,300 MB. When taking into account the amount of time required to load, the MLP required only 0.35 seconds. However, when it came to making a forecast, the two algorithms that were the quickest were the LR and the SGDClassifier. Both of these desired only 0.06 seconds to make a prediction. In spite of the fact that the MLP classifiers had the smallest design sizes and the quickest loading times, the LR was still capable of achieving a higher accuracy score. On the other hand, because of this, MLP classifiers are preferred for a decision-maker that places a higher priority on the size and the amount of time than they do on the accuracy.

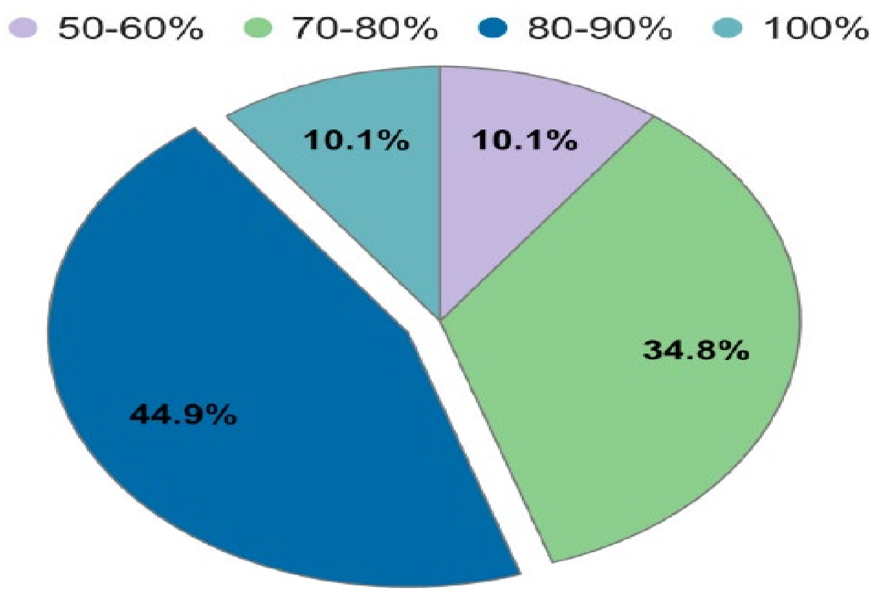

Fig.7. Qualitative investigation dependent on expert doctors.

Figure 7 is a pie chart showing the qualitative assessment of the proposed component. According to the data, over half (44.9%) of the anticipated diagnose are accurate to within (80-90)%, while nearly a third (34.8%) are accurate to within (70-80)%. Additionally, the first 10% is correct between (50-60)% of the time, and the final 10% is correct 100% of the time. The results of the suggested module's quantitative analysis agree strikingly with those of the experts' qualitative analysis, demonstrating the model's sturdiness and the reliability of projected diagnoses. Another example: one of Dr. Altibbi's patients came in with a runny nose and needed an appointment. Two applicable diagnoses and two irrelevant diagnoses were indicated by the created model. In the meantime, the doctor decided that a simple cold was to blame. According to Altibbi's doctors' qualitative assessment, however, the established QSDM model has certain drawbacks. For one, the indicated diagnosis may include duplication on occasion, such as diagnosing the common cold twice. Second, the model may miss the possibility that a patient's symptoms are caused by a fairly common disease. The common cold is an example of a symptom that might not be considered at first. These constraints could slow down the diagnostic process or prevent clinicians from making the best decision possible. Consequently, addressing these constraints is crucial for enhancing the established QSDM model.

Table 1. Performance of models

	Acc			
	R_I	R_II	Sum	Multiply
P_1	91.3	83.2	93.5	94.9
P_2	73.2	81.7	79.2	83.2
P_3	76.1	76.5	77.	79.6

In Table 1 and Figure 8, we see how well the final fused models perform with respect to 4 fusion criteria (R-I, R-II, Sum, &Multiply), as measured by correctness scores for anticipating 20%, 40%, and 80% of the diagnose (depicted by P_1, P_2, and P_3, respectively). It is evident that P_1 produced the highest degree of precision. Multiplication-based fusion, on the other hand, achieved a 94.9% accuracy; this was followed by summation's 84.6%; then R-I's 92.8%; and R-II's 91.3%. There is a distinct and significant difference between P_3 and P_1, despite the fact that P_2 and P_3 behave similarly.

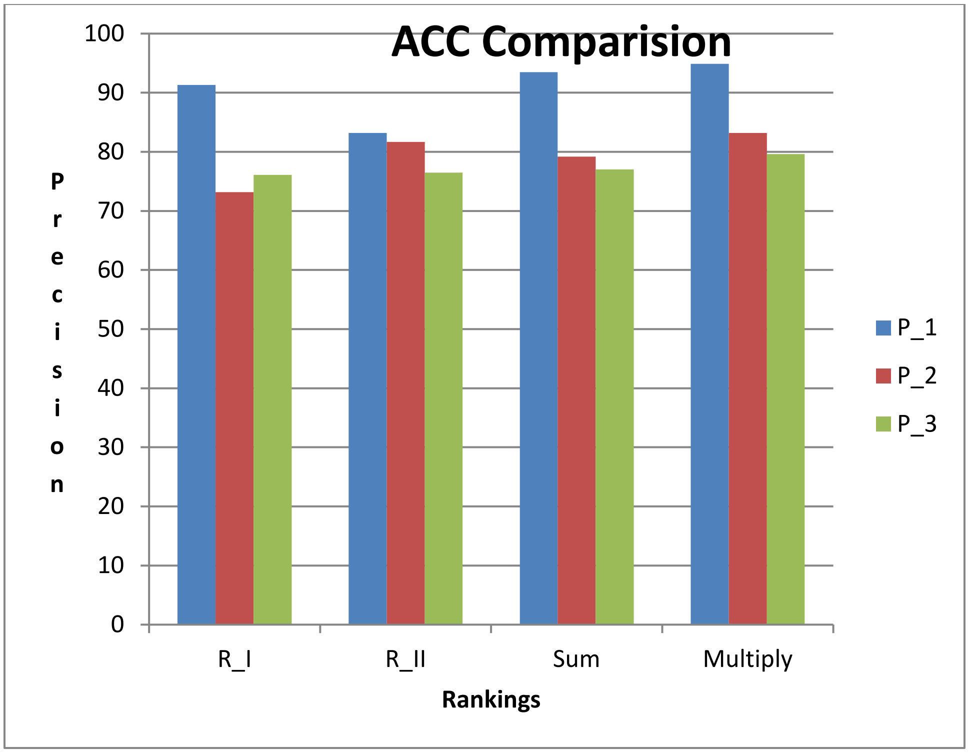

Fig. 8. Performance comparison of models

4.1 Challenges of the Multimodel DL in healthcare

The use of data, models, and the completion of complex tasks all provide difficulties that must be overcome for multi-modal DL to proceed successfully in the area of medicine [42]. The primary obstacles are outlined below.

1. The heterogeneity and diversity of participants, the sample size, the depth of phenotypic analysis, the level of data consistency and synchronization, and the scale of association amid data sources all combine to create a challenge that is greater than that posed by a single-mode DLM in the medical field. Therefore, it is important to take into account the difficulty of working with the vastly varied data seen in real-world medical databases.

2. The gathering, connecting, and cost-effective footnote of multi-dimensional medical-data presents, problems withrespect to cost and speed in the development of multi-modal medical research and clinical applications.

3. Good feature extraction and accurate data type association are prerequisites for multi-modal medical fusion. However, there are several obstacles to overcome, such as the prevalence of small and partial datasets and non-standardized data structures.

4. In some domains, such as three-dimensional imaging &genomics, dispensation even a single instant of data necessitates a substantial total of computational capacity. Therefore, it is a key problem to construct models that can rapidly and simultaneously process massive amounts of data such as tumor pathology slides, genomes, or medical text.

5. Patients and doctors may be concerned about their privacy when their health and clinical data is collected for study. It is crucial to set up a reliable process to monitor and address these problems, and this calls on researchers to propose and investigate further solutions.

6. Clinicians and doctors need to collaborate frequently to design research schemes in the field of multi-modal medical data fusion psychoanalysis. There are many barriers to working together that make this kind of communication difficult.

The medical community and AI researchers will need to work together to build and validate new models and ultimately demonstrate their ability to improve diagnosis and treatment. The use of data, models, and the completion of complex tasks all provide difficulties that must be overcome for multi-modal DL to proceed successfully in the field of medicine [43]. The primary obstacles are outlined below.

1. Problem is that there aren't universally accepted standards for collecting and annotating data, which might introduce bias and reduce model generalizability.

2. It is currently difficult to interpret the results of multi-modal deep-learning models, which could limit their use in clinical settings.

3. Data heterogeneity across modalities, which may necessitate a variety of preprocessing and integration strategies.

4. The accessibility and availability of multi-modal data, as the collection and integration of data from numerous sources may necessitate cooperation between different institutions and data sharing agreements.

5. Model overfitting, which is especially troublesome in medical applications because it can escort to pitiable performance on new data.

Impact on patient privacy and autonomy, as well as other potential ethical concerns related to the use of deep-learning models in medical decision-making.

Few Visions

1. The capacity to generate more accurate and tailored predictions using MMDL shows significant potential for enhancing medical investigation and treatment.

2. Multi-modal data integration can aid in the accurate scrutiny and prediction of diseases like cancer &Alzheimer's, as well as in the prediction of disease risk and the individualization of treatment.

3. MMDL can improve clinical decision-making by providing more precise and quicker diagnoses and treatment suggestions.

Finally, MMDL shows considerable promise for enhancing healthcare outcomes and individualized treatment when applied in medical research and clinical practice. However, there are problems with these models that must be solved. These include things like data heterogeneity, interpretability, and ethical concerns. Together, we can overcome these obstacles and realize the full promise of MMDL in healthcare, leading to better patient diagnosis, care, and outcomes.

V. CONCLUSIONS AND FUTURE WORKS

Early on in the progression of a disease, symptoms are often vague and can easily be mistaken for those of another condition, making it difficult to provide a previous, correct differential diagnosis. It is crucial to create a computer-aided diagnosis system that would aid doctors in building accurate diagnoses. Differential scrutiny decisions made during consultations at Altibbi can be aided by the MMDL investigative system suggested in this study. The proposed method integrates the use of symptoms and queries. Differential diagnosis between the two modalities has been accomplished using a wide variety of machine learning methods. There have been a number of feature extraction techniques used in the questions module, including term frequency, a hashing vectorizer, and document embeddings. The final model is the result of a late fusion of two models; this fusion is accomplished in a number of different ways, including via ranking, summing, and multiplying. The best results in terms of accuracy (84.9%) were found using the fusion method that relied on multiplication. As a result, this design may hold promise as the basis for a differential diagnosis-capable decision-support system. However, boosting the model's precision is crucial. The growing number of Altibbi consultations is a great commodity that may be used to improve the model's effectiveness. In addition, this causes the structural symptoms to worsen. In order to improve accuracy, modern computing techniques like deep learning and Transformers approaches can be applied to enormous amounts of data. The model's weaknesses can be addressed and improved upon by incorporating a third modality, such as diagnostic test and laboratory findings, into the categorization process.

Beyond the scope of this article, the potential uses of multimodal medical AI in healthcare are vast. Many activities in the realm of drug discovery, such as target classification and confirmation, anticipation of drug interactions, and prediction of adverse effects177, could benefit from the use of multidimensional data. Although we covered a lot of ground in this analysis, there are still significant obstacles to the widespread adoption of multimodal AI, such as the possibility of false-positives and how physicians should evaluate and convey the dangers to patients.

REFERENCES

1. Faris, H., Habib, M., Faris, M., Elayan, H. and Alomari, A., 2021. An intelligent multimodal medical diagnosis system based on patients' medical questions and structured symptoms for telemedicine. *Informatics in Medicine Unlocked*, *23*, p.100513.
2. Lim, A., Hoek, H.W. and Blom, J.D., 2015. The attribution of psychotic symptoms to jinn in Islamic patients. *Transcultural psychiatry*, *52*(1), pp.18-32.
3. Acosta, J.N., Falcone, G.J., Rajpurkar, P. and Topol, E.J., 2022. Multimodal biomedical AI. *Nature Medicine*, *28*(9), pp.1773-1784.
4. Reddy, P.C.S., Pradeepa, M., Venkatakiran, S., Walia, R. and Saravanan, M., 2021. Image and signal processing in the underwater environment. *J Nucl Ene Sci Power Generat Techno*, *10*(9), p.2.

5. Alberdi, A., Aztiria, A. and Basarab, A., 2016. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial intelligence in medicine*, *71*, pp.1-29.
6. Stanley, A. and Kucera, J., 2021. Smart healthcare devices and applications, machine learning-based automated diagnostic systems, and real-time medical data analytics in COVID-19 screening, testing, and treatment. *American Journal of Medical Research*, *8*(2), pp.105-117.
7. Villa-Parra, A.C., Criollo, I., Valadão, C., Silva, L., Coelho, Y., Lampier, L., Rangel, L., Sharma, G., Delisle-Rodríguez, D., Calle-Siguencia, J. and Urgiles-Ortiz, F., 2022. Towards multimodal equipment to help in the diagnosis of COVID-19 using machine learning algorithms. *Sensors*, *22*(12), p.4341.
8. A. K. Maurya, K. Lokesh, Sandeep, R. Kumar and R. Krishnamoorthy, 2022. Deep Neuro-Fuzzy Logic Technique for Brain Meningiomasa Prediction, *7th International Conference on Communication and Electronics Systems (ICCES),* pp. 1244-1248, doi: 10.1109/ICCES54183.2022.9836008.
9. Shams, A.B., Raihan, M., Khan, M., Preo, R. and Monjur, O., 2021. Telehealthcare and Covid-19: A Noninvasive & Low Cost Invasive, Scalable and Multimodal Real-Time Smartphone Application for Early Diagnosis of SARS-CoV-2 Infection. *arXiv preprint arXiv:2109.07846*.
10. Ammar, N., Zareie, P., Hare, M.E., Rogers, L., Madubuonwu, S., Yaun, J. and Shaban-Nejad, A., 2021, December. SPACES: Explainable multimodal ai for active surveillance, diagnosis, and management of adverse childhood experiences (ACEs). In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 5843-5847). IEEE.
11. G. N. Vivekananda, A. R. H. Ali, S. Arun, P. Mishra, R. Sengar and R. Krishnamoorthy, 2022. Cloud Based Effective Health Care Management System With Artificial Intelligence," *IEEE 7th International conference for Convergence in Technology (I2CT)*, pp. 1-6, doi: 10.1109/I2CT54291.2022.9825457.
12. Cai, Q., Wang, H., Li, Z. and Liu, X., 2019. A survey on multimodal data-driven smart healthcare systems: approaches and applications. *IEEE Access*, *7*, pp.133583-133599.
13. Reddy, P.C.S., Sucharitha, Y. and Narayana, G.S., 2021. Forecasting of Covid-19 Virus Spread Using Machine Learning Algorithm. *International Journal of Biology and Biomedicine*, *6*.

14. Guohou, S., Lina, Z. and Dongsong, Z., 2020. What reveals about depression level? The role of multimodal features at the level of interview questions. *Information & Management*, *57*(7), p.103349.
15. Shanmugaraja, P., Bhardwaj, M., Mehbodniya, A., VALI, S. and Reddy, P.C.S., 2023. An Efficient Clustered M-path Sinkhole Attack Detection (MSAD) Algorithm for Wireless Sensor Networks. *Adhoc & Sensor Wireless Networks*, *55*.
16. Zheng, G., Zhang, D. and Zhao, W., 2021. Guest Editorial Multi-Modal Computing for Biomedical Intelligence Systems. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), pp.3256-3257.
17. Sucharitha, Y., Reddy, P.C.S. and Suryanarayana, G., 2023. Network Intrusion Detection of Drones Using Recurrent Neural Networks. *Drone Technology: Future Trends and Practical Applications*, pp.375-392.
18. Vásquez-Correa, J.C., Arias-Vergara, T., Orozco-Arroyave, J.R., Eskofier, B., Klucken, J. and Nöth, E., 2018. Multimodal assessment of Parkinson's disease: a deep learning approach. *IEEE journal of biomedical and health informatics*, *23*(4), pp.1618-1630.
19. Dhanalakshmi, R., Bhavani, N.P.G., Raju, S.S., Shaker Reddy, P.C., Marvaluru, D., Singh, D.P. and Batu, A., 2022. Onboard Pointing Error Detection and Estimation of Observation Satellite Data Using Extended Kalman Filter. *Computational Intelligence and Neuroscience*, *2022*.
20. Van Abbema, R., Van Wilgen, C.P., Van Der Schans, C.P. and Van Ittersum, M.W., 2011. Patients with more severe symptoms benefit the most from an intensive multimodal programme in patients with fibromyalgia. *Disability and Rehabilitation*, *33*(9), pp.743-750.
21. Rahman, T., Ibtehaz, N., Khandakar, A., Hossain, M.S.A., Mekki, Y.M.S., Ezeddin, M., Bhuiyan, E.H., Ayari, M.A., Tahir, A., Qiblawey, Y. and Mahmud, S., 2022. QUCoughScope: an intelligent application to detect COVID-19 patients using cough and breath sounds. *Diagnostics*, *12*(4), p.920.
22. B. Singh, K. Somasekhar, K. Anand, M. Gopikrishnan and R. Krishnamoorthy, 2022. Machine learning based Predictive Modeling of Plasma Treatment in Biomedical Surfaces, *Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pp. 1043-1046, doi: 10.1109/ICAIS53314.2022.9743031.

23. Wang, C., Wang, H., Zhuang, H., Li, W., Han, S., Zhang, H. and Zhuang, L., 2020. Chinese medical named entity recognition based on multi-granularity semantic dictionary and multimodal tree. *Journal of Biomedical Informatics*, *111*, p.103583.
24. Prasath, A.S.S., Lokesh, S., Krishnakumar, N.J., Vandarkuzhali, T., Sahu, D.N. and Reddy, P.C.S., 2022. Classification of EEG signals using machine learning and deep learning techniques. *International journal of health sciences*, *2022*, pp.10794-10807.
25. Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F. and Luo, Y., 2022. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, *5*(1), p.171.
26. Muthappa, K.A., Nisha, A.S.A., Shastri, R., Avasthi, V. and Reddy, P.C.S., 2023. Design of high-speed, low-power non-volatile master slave flip flop (NVMSFF) for memory registers designs. *Applied Nanoscience*, pp.1-10.
27. Al Bassam, N., Hussain, S.A., Al Qaraghuli, A., Khan, J., Sumesh, E.P. and Lavanya, V., 2021. IoT based wearable device to monitor the signs of quarantined remote patients of COVID-19. *Informatics in medicine unlocked*, *24*, p.100588.
28. Cavalcanti, T.C., Lew, H.M., Lee, K., Lee, S.Y., Park, M.K. and Hwang, J.Y., 2021. Intelligent smartphone-based multimode imaging otoscope for the mobile diagnosis of otitis media. *Biomedical Optics Express*, *12*(12), pp.7765-7779.
29. Sabitha, R., Shukla, A.P., Mehbodniya, A., Shakkeera, L. and REDDY, P.C.S., 2022. A Fuzzy Trust Evaluation of Cloud Collaboration Outlier Detection in Wireless Sensor Networks. *Adhoc & Sensor Wireless Networks*, *53*.
30. Yang, G., Ye, Q. and Xia, J., 2022. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Information Fusion*, *77*, pp.29-52.
31. Ashok, K., Boddu, R., Syed, S.A., Sonawane, V.R., Dabhade, R.G. and Reddy, P.C.S., 2022. GAN Base feedback analysis system for industrial IOT networks. *Automatika*, pp.1-9.
32. Shams, A.B., Raihan, M., Sarker, M., Khan, M., Uddin, M., Monjur, O. and Preo, R.B., 2021. Telehealthcare and Telepathology in Pandemic: A Noninvasive, Low-Cost Micro-Invasive and Multimodal Real-Time Online Application for Early Diagnosis of COVID-19 Infection. *arXiv preprint arXiv:2109.07846*.

33. Shaker Reddy, P.C. and Sucharitha, Y., 2022. IoT-Enabled Energy-efficient Multipath Power Control for Underwater Sensor Networks. *International Journal of Sensors Wireless Communications and Control*, *12*(6), pp.478-494.
34. Roy, S., Meena, T. and Lim, S.J., 2022. Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. *Diagnostics*, *12*(10), p.2549.
35. Lokesh, S., Priya, A., Sakhare, D.T., Devi, R.M., Sahu, D.N. and Reddy, P.C.S., 2022. CNN based deep learning methods for precise analysis of cardiac arrhythmias. *International journal of health sciences*, *6*.
36. Gyrard, A., Jaimini, U., Gaur, M., Shekharpour, S., Thirunarayan, K. and Sheth, A., 2022. Reasoning over personalized healthcare knowledge graph: a case study of patients with allergies and symptoms. In *Semantic Models in IoT and Ehealth Applications* (pp. 199-225). Academic Press.
37. Kumar, K., Pande, S.V., Kumar, T., Saini, P., Chaturvedi, A., Reddy, P.C.S. and Shah, K.B., 2023. Intelligent controller design and fault prediction using machine learning model. *International Transactions on Electrical Energy Systems*, *2023*.
38. Song, W., Hou, X., Li, S., Chen, C., Gao, D., Sun, Y., Hou, J. and Hao, A., 2022. An Intelligent Virtual Standard Patient for Medical Students Training Based on Oral Knowledge Graph. *IEEE Transactions on Multimedia*.
39. Ashreetha, B., Devi, M.R., Kumar, U.P., Mani, M.K., Sahu, D.N. and Reddy, P.C.S., 2022. Soft optimization techniques for automatic liver cancer detection in abdominal liver images. *International journal of health sciences*, *6*.
40. Liu, L., Shafiq, M., Sonawane, V.R., Murthy, M.Y.B., Reddy, P.C.S. and kumar Reddy, K.C., 2022. Spectrum trading and sharing in unmanned aerial vehicles based on distributed blockchain consortium system. *Computers and Electrical Engineering*, *103*, p.108255.
41. Faris, H., Faris, M., Habib, M. and Alomari, A., 2022. Automatic symptoms identification from a massive volume of unstructured medical consultations using deep neural and BERT models. *Heliyon*, *8*(6), p.e09683.
42. Reddy, P.C.S., Suryanarayana, G. and Yadala, S., 2022, November. Data analytics in farming: rice price prediction in Andhra Pradesh. In *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)* (pp. 1-5). IEEE.

43. Palliya Guruge, C., Oviatt, S., Delir Haghighi, P. and Pritchard, E., 2021, October. Advances in multimodal behavioral analytics for early dementia diagnosis: a review. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (pp. 328-340).