IJICIS, Vol.20, No. 1

International Journal of Intelligent Computing and Information Science

A Medical Probabilistic Advisory System (MPAS) Based on Independent Artificial Intelligent Techniques to Support Decision-Making

Ahmed E. Amin

Department of Computer Science,
Mansoura University, Mansoura
35516, Egypt
ahmedel_sayed@mans.edu.eg

Abstract:
In this paper, an intelligent advisory system is used to make the decision depend on extracting the features of events by using artificial intelligence techniques. The proposed system was applied in the training in medical institutions in the field of physical therapy. X-ray images and their corresponding reports are used as events. For X-ray images, the Co-occurrence matrix is used to extract the features and Evidential K-Nearest Neighboring Rule has used to features classification then stored in the database. Whereas, the hidden Markov model is used to recognize the handwritten reports after their processing. The probability theory is used to solve the feature's overlapping problem that makes it easier to knowledge base formation by experts.

The proposed system has been applied to different and several pathological patients of bone fractures and jaw and dental injuries. Comparing the results with the experts has shown the high accuracy that characterizes the proposed system.

Keywords: decision support systems, advisory systems, rule based systems, probabilistic theory, Hidden Markova Model.

1. Introduction

Since a long time, computer scientists are interested in a decision-making process, which is considered a form of reasoning towards action. The decision-making contributes to choosing the best action among the various alternatives that are available to have some information about the current state and the consequences of potential actions [1]. An advisory system can be classified as a type of expert systems [2, 3] which provide the advices and assist in solving problems that are normally solved by human experts. Experts are employing some of the mathematical methods that rely on probabilistic treatment of uncertainty in various areas such as engineering, science, management and medicine. These methods have proven over time to lead to results of predictable reliability and have become integral parts of both teaching and practice in each of the fields. It seems natural to base probabilistic advisory systems (PAS) in these domains on the normative foundations of probability theory and decision theory. PAS are constructed by eliciting knowledge from human experts and coding it into a form that can be used by a computer in the evaluation of alternative solutions to problems within that domain of expertise.

The decision-maker is collaborating with the advisory system that identifies the elements of the problem from input data processing (such as images and text) and extracts features of each element separately then evaluates possible solutions to derive optimal decision. For example, a medical advisory system can be used to help determine physiotherapy system after undergoing surgery, based on the processing of X-ray images and its corresponding reports. In this example, physiotherapist is legally responsible for the decision to determine the treatment regimen.

Advisory systems that used to support decisions can be classified as intelligent or unstructured. Both of advisory system are characterized by novelty, complexity, and open ended-ness [4]. In addition to these characteristics, contextual uncertainty is ubiquitous in unstructured decisions, which when combined exponentially increases the complexity of the decision-making process. As a result of the previous characteristics and uncertain solutions, theories of probability have been used with knowledge bases [5] to evaluate alternative paths to determine the optimal path that has the highest probability of a desirable outcome [6]. Fortunately, the medical field is considered as a fertile ground for building knowledge-based systems that can serve as high-level advisory [7].

The paper is organized as follows: Section 2 Basic Notation. Section 3 image and text evidences. Section 4 is devoted to fusion of evidence and decision making. Section 5 proposed method and Section 6 presents experimental and results. The paper is terminated by concluding and summarizing the obtained results.

2. Basic Notation:

A. Basics of Dempster-Shafer Theory

In Dempster-Shafer (DS) theory [8], a frame of discernment Ω is defined as the set of all hypotheses in a certain domain. A basic belief assignment (BBA) is a function *m* that defines the mapping from the power set of Ω to the interval [0; 1] and verifies:

$$m: 2^{\Omega} \longrightarrow [0,1] \quad \dots\dots\dots\dots \quad (1)$$

$$\sum_{A \in 2^{\Omega}} m(A) = 1 \quad \dots\dots\dots\dots \quad (2)$$

The quantity $m(A)$ can be interpreted as a measure of the belief that is committed exactly to *A*, given the available evidence. A subset $A \in 2^{\Omega}$ with $m(A) > 0$ is called a focal element of *m*. In DS theory, two functions of evidence can be deduced from *m* and its associated focal elements, belief function Bel and plausibility function Pl. $Bel(A)$ is the measure of the total belief committed to a set *A*. The belief function is defined as a mapping $Bel : 2^{\Omega} \rightarrow [01]$ that satisfies $Bel(\phi) = 0$, $Bel(\Omega) = 1$ and for each focal element A, we have:

$$Bel(A) = \sum_{\phi \neq B \subseteq A} m(B) \qquad \dots\dots\dots\dots\dots \qquad (3)$$

The plausibility of *A*, $Pl(A)$, represents the amounts of belief that could potentially place in *A* and defined as:

$$Pl(A) = \sum_{A \cap B \neq \phi} m(B) \qquad \dots\dots\dots\dots\dots \qquad (4)$$

- **Dempster's Combination Rule**

When there are many sources of information defined on the same frame of discernment, the mass functions from different sources are combined under the normalized Dempster's combination rule [8].

$$m_{12}(A) = m_1 \oplus m_2 = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \phi} m_1(B) m_2(C)} , & \forall A \subseteq \Omega \, , \, A \neq \phi \\ 0 & if \;\; A = \phi \end{cases} \qquad \dots\dots \qquad (5)$$

Where: $k = \sum_{B \cap C = \phi} m_1(B) m_2(C)$ represents the degree of conflict between the two sources. If $k = 1$ the two evidences are in conflict and they cannot be combined.

- **Pignistic probability Transformation:**

Having combined all the available evidences, we always turn to the problem of making decisions. To do this, one simple approach to convert a belief function to a probability function is often adopted, that is Pignistic Probability Transformation $(BelP)$ [9].

$$BelP(B) = \sum_{A \in \theta} m(A) \frac{|B \cap A|}{|A|} \quad ...... \quad (6)$$

Where $|A|$ denotes the number of elements of θ in A.

- **Evidence Theory:**

Evidence theory allows using uncertain data [10]. Let Ω be a finite set of mutually exclusive and exhaustive hypotheses, called the frame of discernment. A basic belief assignment is a function m from 2Ω to [0, 1] verifying:

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \subseteq \Omega} m(\emptyset) = 1 \end{cases}$$

For any $A \subseteq \Omega$, $m(A)$ represents the belief that one is willing to commit exactly to A, given a certain piece of evidence. The subsets (A) from Ω $(m(A) > 0)$ are called the focal elements of m. Associated with m are a belief or credibility function bel and a plausibility function pl, defined, respectively, for all $A \in \Omega$ as:

$$bel(A) = \sum_{B \subseteq A} m(B)$$

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$$

The quantity bel(A) can be interpreted as a global measure of one's belief that hypothesis is true, while pl(A) may be viewed as the amount of belief that could potentially be placed in A, if further information became available [10]. The decision rule can be given by different approaches as following:

- Choose the maximum plausibility hypothesis (pl);
- Choose the maximum Pignistique probability hypothesis $(BetP)$.

$$\mathrm{BetP(w)} = \sum_{\mathrm{w \in A}} \frac{\mathrm{m(A)}}{\mathrm{|A|}}$$

Evidence representing algorithm [10]:
1- Entries generation: **For each $i \in [1, n]$, generates $\Phi \in 2^{\Phi}$ according to the basic belief function (m_i), considered as probabilistic distribution over the set 2^{Φ}.** **2- Conditional arbitrament:** ❖ **Generate $\varphi \in 2^{\Phi}$ according to experts $E(\varphi \mid \Phi_{1:n}; m_{1:n})$ considered as a probabilistic distribution over the set 2^{Φ}.** ❖ **In case empty set $\Phi = \phi$ means impossible event. Otherwise event is focal element.**

B. Image and Text Evidences:

I. Image's Feature Extraction Technique:

The complex information embedded in image is searched to restore into vectors defined as features. The feature extraction process is depending on image type which classified into binary, grayscale, indexed and true color. The important feature of many image types is texture [11] that the pattern of information of structure found in an image. Texture features can be extracted in several methods, using statistical [12], structural [13], model-based and transform information [14], in which the most common way is using the Gray Level Co-occurrence Matrix (GLCM).

- **Co-occurrence Matrix (CM):**

Co-occurrence matrix $\left(CM_{d\theta}(i,j)\right)$ is defined as the second-order of image histogram for each distance *d* and orientation *θ*. The given image is represented by $f(x,y)$ with a set of *G* discrete intensity levels. The matrix $CM_{d\theta}(i,j)$ is defined as its $(i,j)^{th}$ entry is equal to the number of times that:

$$f(x_1,y_1)=i \quad and \quad f(x_2,y_2)=j$$

where

$$(x_2,y_2)=(x_1,y_1)+(d\cos\theta, d\sin\theta)$$

There are two forms of CM, one symmetric [15] where pairs separated by $\pm d$ for a given direction θ are counted, the other non-symmetric [16] where only pairs separated by distance *d* are counted.

The joint between two pixels probability $p_{d\theta}(i,j)$ is estimated by divided the elements of CM by the total number of neighboring pixels $\Re(d,\theta)$ in the image [17-18].

- **Feature extracting based on CM:**

According to co-occurrence matrix, Haralick [19] defines fourteen textural features measured from the probability matrix to extract the characteristics of texture statistics. Some of the texture features which contain information about image textural characteristics are defined by:

$$Energy(E)=\sum_{i=0}^{x-1}\sum_{j=0}^{y-1}\left[p(i,j)\right]^2 \quad .... \quad (7)$$

$$Contrast(C)=\sum_{i=0}^{x-1}\sum_{j=0}^{y-1}(i-j)^2\, p(i,j) \quad ..... \quad (8)$$

$$Homogeneity(H)=\sum_{i=0}^{x-1}\sum_{j=0}^{y-1}\frac{p(i,j)}{1+(i,j)^2} \quad ...... \quad (9)$$

$$Correlation(Co)=\sum_{i=0}^{x-1}\sum_{j=0}^{y-1}\frac{ijp(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y} \quad ..... \quad (10)$$

Where μ_x, μ_y, σ_x *and* σ_y denoted as the mean and standard deviations of the row and column sums of the matrix respectively.

The collection of all such features forms the feature vector. A feature vector is given by:

$$f_{im_i}^{CL_j}=\{E_\theta, C_\theta, H_\theta, Co_\theta\}$$

Where, i: image number, j: cluster number, *θ*: four possible directions ($0^\circ, 45^\circ, 90^\circ$ and 135°).

- **Classification of features:**

Image classification aims to detection, identification, and classification of features which occurring in an image class type, these features represent on the field.Classes overlapping are a common problem in the image classification; it occurs when there are ambiguous regions in the class where the prior probability for both classes is approximately equal. Therefore, feature classification can be used Dempster-Shafer theory (DST) of an evidence-based approach. DST is characterized by the capability to handle the uncertainties that might arise due to the similarity of different expression features, and distributions of these features for different expression may overlap. On the other hand, the mass values are calculated by DST that represented by Evidential K-Nearest Neighboring Rule.

- **Evidential K-Nearest Neighboring Rule (EKNNR):**

The mass values are calculated from the EKNNR patterns according to a distance metric for a given query pattern. The evidences from the neighbors are then combined using DS rule of combination.
Let $U = \{CL_1, CL_2, ..., CL_n\}$ denoted as different image classes; Q is the feature vector to be classified and (N_K) is the set of its k-nearest neighbors in training set. Any $Q_i \in N_K$ may belong to a class $CL_q \in U$. This membership of Q_i can provide piece of evidence to increase our belief that Q belong to CL_q. The strength of this evidence decreases with the distance d_i between Q and Q_i. This evidence is represented by mass value m_i as follows:

$$m_i(\{CL_q\}) = \varphi(di)$$
$$m_i(U) = 1 - \varphi(d_i)$$

Where $\varphi(.)$ is a decreasing function of Euclidean distance from $[0,+\infty)$ to $[0,1]$ such that $\lim_{d \to +\infty} \varphi(d) = 0$:

$$\varphi(d_i) = \alpha \exp\left(-\gamma_q^2 (d_i)^2\right)$$

Where α is constant and γ_q is a parameter associated to class CL_q. Such pieces of KNN evidences are pooled using Dempater's rule of combination:

$$m = \bigoplus_{Q_i \in N_K} m_i = \bigoplus_{Q_i \in N_K} \{Q_i\}^{1-\varphi(d_i)}$$

And Q is classified to the class, for which the Pignistic probability [20] is maximum.

II. Text Evidence:

Independently, an X-ray image' report which was handwritten can be produced by a machine. In machine print, the form of the individual symbols or characters in principle is not constrained. In the text recognition, it is necessary to the analysis of the report layout. Where there are not only available page image text areas but also other elements of the report structure such as headlines or graphics, are identified. The text areas can be segmented into paragraphs, individual lines, and in general, also single characters due to the usually high precision in the production of machine-printed reports.
As soon as the images of the written symbols are isolated, they can be mapped onto a symbolic representation by arbitrary techniques from pattern classification. The results of the classification are generally subject to one or more post-processing steps. There it is tried, to correct errors on the character level as far as possible by incorporating context restrictions, e.g. in the form of a lexicon.

- **Handwriting Markov Model:**

The discrete models which using to handwritten report recognition, they require the use of vector quantizer, which converts continuous feature representations to discrete observation sequences prior to the analysis. In order to be able to nevertheless deal with arbitrary continuous distributions with multiple modes or regions of high density in general, approximatively techniques are applied. The most widely used method consists in the use of mixture densities on the basis of Gaussian densities. In general, every general continuous probability distribution $p(x)$ can be approximated to arbitrary precision with a linear combination of infinitely many component normal distributions:

$$p(x) \cong \sum_{k=1}^{\infty} c_k \mathcal{N}(x|\mu_k, C_k) \approx \sum_{k=1}^{M} c_k \mathcal{N}(x|\mu_k, C_k)$$
$$\sum_k c_k = 1 \quad with \;\; 0 \le c_k \le 1 \;\; \forall k$$

Where M is a finite number of mixtures.

The general form of continuous HMMs, therefore, uses one mixture density per state j for the description of the output probability density function:

$$b_j(x) = \sum_{k=1}^{M_j} c_{jk}\mathcal{N}\left(x \middle| \mu_{jk}, C_{jk}\right) = \sum_{k=1}^{M_j} c_{jk} g_{jk}(x)$$

Where M_j a number of mixture components which is used may vary from state to state. $g_{jk}(x)$ denoted to each of the normal distributions, furthermore possesses an individual set of parameters consisting of a mean vector μ_{jk} and a covariance matrix C_{jk}.

Semi-continuous HMMs approaches are developed which are frequently referred to as tied mixture models. In such models only a single set of component densities is used to construct all state-specific output probability densities:

$$b_j(x) = \sum_{k=1}^{M} c_{jk}\mathcal{N}(x|\mu_k, C_k) = \sum_{k=1}^{M} c_{jk} g_k(x)$$

In semi-continuous HMMs every mixture density consists of the same number M of baseline distributions.

- **Handwrite Report Recognition:**

In most approaches for handwriting recognition, which consider isolated phrases, e.g., in X-Ray image report, it is the goal of the system described in the following:

Preprocessing → Feature Extraction → Hand writing Model

- **Preprocessing:**

In handwritten reports (HwR), normalization operations are used not only for extracting features of the words in order to compensate for the letter's tendency that is represented by drift and slant on baseline but also variations with in a line can be captured approximately [21]. Then a local binarization of the text line image is performed. Thus it is ensured, that intensity variations of both writing and background do not adversely affect the subsequent feature extraction. As a final preprocessing step, the text line image is normalized in size. For this purpose first local extrema of the contour of the writing are determined. Then the line image is rescaled such, that the average distance between these matches a predefined constant.

- **Feature Extraction:**

Sliding window technique is used to determine the handwritten reports features. This technique is used to convert the text line images into sequences of feature vectors where they are subdivided text line images after pre-segmented and normalized to the analysis windows, which are four pixels wide and overlap each other by half. For each of these windows, nine geometric features are computed can be classified into groups. The first group of features describes the coarse shape of the writing the local analysis window which computes the average distance of the lower baseline to both the upper and the lower contour, and the distance of the center of gravity of text pixels to the baseline. Then these features are normalized by the core size in order to increase the robustness against variations in the size of the writing. The second group are three local trend attributes calculated. The second group, three local directional features are calculated describing the orientation of the lower and upper contour as well as the gradient of the mean of the column-wise pixel distributions. Finally, the average number of black-to-white transitions per column, the average number of text pixels per column, and the average number of text pixels between upper and lower contour are calculated.

- **Handwriting Model:**

Semi-continuous hidden Markova model (SHMM) that the statistical modeling is used to represent the handwriting report model [22]. SHMM is performed with a shared code of approximately 2000 Gaussians with diagonal covariance's. Modeling for 52 letters, 10 digits, 12 punctuation symbols, and white space are created by a total of 75 context independent SHMM. The number of model states is automatically determined depending on the length of the respective unit in the training material. All these models use the Bakis topology [23] in order to be able to capture a wider variability in the length of the character patterns described.

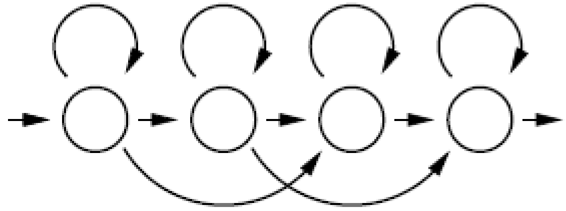


Fig. 1: Bakis topology [23]

C. Probabilistic Evidence Classification System:

For each X-ray image and corresponding handwritten report features database which consist of gathering of image's features called clusters. The *n* set of clusters can be defined as the frame of discernment of image clusters $U = \{CL_1, CL_2, ..., CL_n\}$. Each cluster is gathering *j* subset of images denoted as $CL_i = \{\mathrm{Im}_1^{CL_i}, \mathrm{Im}_2^{CL_i}, ..., \mathrm{Im}_j^{CL_i}\}$. Feature extraction for each image $\left(F_{im_j}^{CL_i}\right)$ is extracted and stored in database namely image features as:

$$F_{im_j}^{CL_i} = \left[\begin{array}{cccc|c} f_{im_{11}}^{CL_1} & f_{im_{12}}^{CL_1} & \cdots & f_{im_{1k}}^{CL_1} & L_{im_{1k}}^{CL_1} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ f_{im_{11}}^{CL_2} & f_{im_{12}}^{CL_2} & \cdots & f_{im_{1k}}^{CL_2} & L_{im_{1k}}^{CL_2} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ f_{im_{11}}^{CL_n} & f_{im_{12}}^{CL_n} & \cdots & f_{im_{1k}}^{CL_n} & L_{im_{1k}}^{CL_n} \end{array}\right],$$

$$where:$$

$$k \text{ is length of image feature}$$

Clusters of image features are divided to intervals, $\left[f_{LO}^{CL_i}, f_{UP}^{CL_i}\right]$ be its lower and upper boundary where $0 \le f_{LO}^{CL_i} \le f_{UP}^{CL_i} \le 1 \ (i = 1,2,...,n)$. An interval-valued belief structure is the belief structure on $\left(F_{im_j}^{CL_i}\right)$ as:

$$f_{LO}^{CL_i} \le m\left(F_{im_j}^{CL_i}\right) \le f_{UP}^{CL_i}$$

$$where:$$

$$0 \le f_{LO}^{CL_i} \le f_{UP}^{CL_i} \le 1 \qquad ..... \ (11)$$

$$for$$

$$i = 1,2,..,n \text{ and } j = 1,2,...,k$$

$$\sum_{i=1}^{m} f_{LO}^{CL_i} \le 1 \text{ and } \sum_{i=1}^{m} f_{UP}^{CL_i} \ge 1 \quad ...... \ (12)$$

$$m\left(F_{im_j}^{CL_i}\right) = 0, \forall F_{im_j}^{CL_i} \notin \{\mathrm{Im}_1^{CL_i}, \mathrm{Im}_2^{CL_i}, ..., \mathrm{Im}_j^{CL_i}\} \qquad ..... \ (13)$$

Interval-valued belief structure (*m*) is normalized with interval-valued probability masses $f_{LO}^{CL_i} \le m\left(F_{im_j}^{CL_i}\right) \le f_{UP}^{CL_i}$ for $i = 1,2...,n$ and $j = 1,2,...,k$. The belief measure (Bel) and plausibility measure (pl) of image feature $\left(F_{im_j}^{CL_i}\right)$ are both the closed intervals defined respectively by:

$$Bel\left(F_{im_j}^{CL_i}\right) = \left[Bel_{LO}\left(F_{im_j}^{CL_i}\right), Bel_{UP}\left(F_{im_j}^{CL_i}\right)\right] \quad ...... \ (14)$$

$$Pl\left(F_{im_j}^{CL_i}\right) = \left[Pl_{LO}\left(F_{im_j}^{CL_i}\right), Pl_{UP}\left(F_{im_j}^{CL_i}\right)\right] \quad ..... \ (15)$$

Where,

$$Bel_{LO}\left(F_{im_j}^{CL_i}\right)=\min\sum_{F_{im_j}^{CL_i}\subseteq CL_i} m\left(F_{im_j}^{CL_i}\right)=\max\left[\sum_{F_{im_j}^{CL_i}\subseteq CL_i} f_{LO}^{CL_i},\left(1-\sum_{F_{im_j}^{CL_i}\not\subset CL_i} f_{UP}^{CL_i}\right)\right] \quad .... \ (16)$$

$$Bel_{UP}\left(F_{im_j}^{CL_i}\right)=\max\sum_{F_{im_j}^{CL_i}\subseteq CL_i} m\left(F_{im_j}^{CL_i}\right)=\min\left[\sum_{F_{im_j}^{CL_i}\subseteq CL_i} f_{UP}^{CL_i},\left(1-\sum_{F_{im_i}^{CL_j}\not\subset F_{im}^{CL}} f_{LO}^{CL_i}\right)\right] \quad .... \ (17)$$

$$pl_{LO}\left(F_{im_j}^{CL_i}\right)=\min\sum_{F_{im_j}^{CL_i}\cap CL_i\neq\phi} m\left(F_{im_j}^{CL_i}\right)=\max\left[\sum_{F_{im_j}^{CL_i}\cap CL_i\neq\phi} f_{LO}^{CL_i},\left(1-\sum_{F_{im_j}^{CL_i}\cap CL_i=\phi} f_{UP}^{CL_i}\right)\right] \quad .... \ (18)$$

$$Pl_{UP}\left(F_{im_j}^{CL_i}\right)=\max\sum_{F_{im_j}^{CL_i}\cap CL_i\neq\phi} m\left(F_{im_j}^{CL_i}\right)=\min\left[\sum_{F_{im_i}^{CL_j}\cap CL_i\neq\phi} f_{UP}^{CL_i},\left(1-\sum_{F_{im_i}^{CL_j}\cap F_{im}^{CL}=\phi} f_{LO}^{CL_i}\right)\right] \quad .... (19)$$

The interval-value for image feature probability is measured by two measures namely feature distance measure $\left(F_{im_j}^{D}\right)$ and feature probability distribution $\left(F_{im_j}^{P}\right)$. The probability of image feature is measured by the most widely measure namely Shannon entropy [24]. This measure can express as:

$$H(p)\cong-\sum_{i=1}^{n} F_{im_j}^{P}\log_2\left(F_{im_j}^{P}\right) \quad ...... \ (20)$$

The probabilistic information content (PIC) [25] for image feature of probability measure $\left(F_{im_j}^{P}\right)$ associated with a probabilistic image feature source over a discrete finite set $F_{im_j}^{P}=\left\{F_{im_1}^{P},F_{im_2}^{P},...,F_{im_j}^{P}\right\}$ is defined by:

$$F_{im_j}^{PIC}=1+\frac{1}{\log_2(n)}\sum_{i=1}^{n} F_{im_j}^{P}\log_2\left(F_{im_j}^{P}\right) \quad ..... \ (21)$$

For each measure the interval structures are defined by:

$$F_{im_j}^{D}=\left[\left(F_{im_j}^{D}\right)_{LO},\left(F_{im_j}^{D}\right)_{UP}\right] \quad and \quad F_{im_j}^{PIC}=\left[\left(F_{im_j}^{PIC}\right)_{LO},\left(F_{im_j}^{PIC}\right)_{UP}\right]$$

Where, the feature image distance $\left(F_{im_j}^{D}\right)$ is the maximum of the following optimization problem:

$$\underset{\left\{F_{f_i}^{P}|F_{im_i}^{CL_j}\in U\right\}}{Max}\left(F_{im_j}^{D}\right)=\left(\sum_{F_{im_i}^{CL_j}\in U}\left(F_{im_j}^{PIC}-m\left(F_{im_j}^{CL_i}\right)\right)^2\right)^{0.5} \quad ..... \ (22)$$

The final feature image distributions $F_{im_j}^{PIC}=\left[\left(F_{im_j}^{PIC}\right)_{LO},\left(F_{im_j}^{PIC}\right)_{UP}\right]$ must meet $F_{im_j}^{PIC}=\left[\min\left(F_{im_j}^{PIC}\right),\max\left(F_{im_j}^{PIC}\right)\right]$ for $j=1,2,...,k$. The constraints of the optimization problem confirm the belief function and probability function respectively as a lower and upper probability. When the interval-valued belief structure is more precise value belief structure at:

$$m\left(F_{im_i}^{CL_j}\right)_{LO}=m\left(F_{im_i}^{CL_j}\right)_{UP}, \ Bel_{LO}\left(F_{im_i}^{CL_j}\right)=Bel_{UP}\left(F_{im_i}^{CL_j}\right), \ Pl_{LO}\left(F_{im_i}^{CL_j}\right)=Pl_{UP}\left(F_{im_i}^{CL_j}\right)$$

Then the constraints of optimality approach are:

$$Const.\begin{cases} Bel\left(F_{im_j}^{CL_i}\right)\leq F_{im_j}^{PIC}\leq Pl\left(F_{im_j}^{CL_i}\right)\rightarrow F_{im_j}^{CL_i}=CL_i \\ Bel\left(F_{im_j}^{CL_i}\right)\leq F_{im_j}^{P}+...+F_{im_j}^{PIC}\leq Pl\left(F_{im_j}^{CL_i}\right)\rightarrow F_{im_j}^{CL_i}=CL_1\cup...\cup CL_n \\ Bel\left(F_{im_j}^{CL_i}\right)\leq\sum_{f_i\in U}F_{im_j}^{PIC}\leq Pl\left(F_{im_j}^{CL_i}\right)\rightarrow F_{im_j}^{CL_i}=U \\ F_{im_j}^{PIC}\leq F_{im_j}^{PIC}\rightarrow if \ \ Pl\left(F_{im_j}^{CL_i}\right)\leq Pl\left(F_{im_j}^{CL_i}\right) \end{cases} \quad ..... \ (23)$$

D. Fusion of Evidence and Decision Making:

- **Evidence Fusion:**

The purpose of aggregation of information is to meaningfully summarize and simplify information rationally obtained from an independent source or multiple sources. Hence the evidence fusion algorithm can be done by algorithm 1. Combination rules are the special types of aggregation methods for data obtained from multiple sources. From a set theoretic standpoint, the combination and disjunction of evidence is employed by AND (set intersection) and OR (set union) operation respectively. The combination rule is determined from the aggregation of two basic probability assignment of m1 and m2 in the following manner:

$$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \qquad where \;\; A \neq \phi \qquad (I)$$

$$m_{12}(\phi) = 0$$

$$K = \sum_{B \cap C = \phi} m_1(B) m_2(C)$$

The normalization factor (1-K) has the effect of completely ignoring conflict and attributing any probability mass associated with conflict to the null set [26]. The combination rule results which based on conjunctive pooled evidence can be measured by evidence measures.

Algorithm 1: Evidence Fusion [26]

Data:

Evidence of Image : $Ev_{\text{Im}} = \{ev_{im_1}, ev_{im_2}, ..., ev_{im_n}\}, \; Ev_{\text{Im}}.bel, \; Ev_{\text{Im}}.focal$

$n - Experts\,(ex): \{ex_1, ex_2 ..., ex_n\}$

$m(ev_{im_k}) = m(ev_{Text_k}) \neq 0$

Evidence of Text: $Ev_{Text} = \{ev_{Text_1}, ev_{Text_2}, ..., ev_{Text_3}\}, \; Ev_{Text}.bel, \; Ev_{Text}.focal$

Results:

Fusion of (Ev_{Im}) **and** (Ev_{Text}): F_{ev}

$for \;\; i = 1{:}n \;\; do$

$fusion \leftarrow \phi$

$for \; Ev_{Text}.focal \;\; in \;\; ex_i \;\; do$

$for \; Ev_{\text{Im}}.focal \;\; in \;\; ex_i \;\; do$

$K \leftarrow Ev_{Text}.focal \cap Ev_{\text{Im}}.focal$

$fusion.focal \leftarrow K$

$fusion.bel \leftarrow Ev_{Text}.bel \times Ev_{\text{Im}}.bel$

$Concatenate \; same \; focal \; in \; fusion$

$F_{ev} \leftarrow fusion$

- **Decision Making:**

A belief function has to be transformed into a probability function for decision making as shown in figure 2. The belief function that quantifies knowledge of the actual class of (x) is transformed into a Pignistic probability distribution [27]. Each mass of belief $m(A)$ is divided equally between the elements of (A) for all $A \subseteq \Phi$. This leads to Pignistic probability distribution of class (w) defined as [28]:

$$BetP(w_k) = \sum_{w_k \in A} \frac{m(A)}{|A|}, \;\; \forall w_k \in \Phi \qquad (II)$$

3. PROPOSED METHOD:

The proposed method presents advisory system for diagnosis and taking the treatment decision based on evidence such as X-ray and its corresponding text report. Figure 3 shown four stages namely, X-ray processing evidence, hand written report processing evidence, fusion of evidences by experts, then building advisory system respectively.

We consider there are two different sources for independent evidence namely image evidence (Ev_{Im}) and corresponding report evidence (Ev_{Text}). Each of evidence detects a set of objects denoted by:

$$\boldsymbol{Ev_{Im_{X-ray}}} = \{\boldsymbol{Ev_{Im_1}}, \boldsymbol{Ev_{Im_2}}, \dots, \boldsymbol{Ev_{Im_n}}\}$$

$$\boldsymbol{Ev_{CTR}} = \{\boldsymbol{Ev_{CTR_1}}, \boldsymbol{Ev_{CTR_2}}, \dots, \boldsymbol{Ev_{CTR_n}}\}$$

All evidence that has been obtained from the classifier that gives information on the actual class of a test pattern. This information can be represented by a belief mass function $m(.)$ after the presentation on the expert. Let $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ be a frame of discernment of a decision making problem under consideration (n) distinct elements φ_i, $i = 1,2, \dots, n$. Evidence belief mass function defined as a mapping from the power set of Φ denoted by 2^{Φ} that must satisfy the two conditions. The first is mass of empty set which represent the impossible event is zero and the other is the mass of belief is normalized to one. An element $\beta \subseteq \Phi$ is called focal element if and only if $m(\beta) > 0$. Focal element represents a degree of belief attached to the proposition $\varphi \in \beta$ and to uncertainly proposition, based on some evidence.

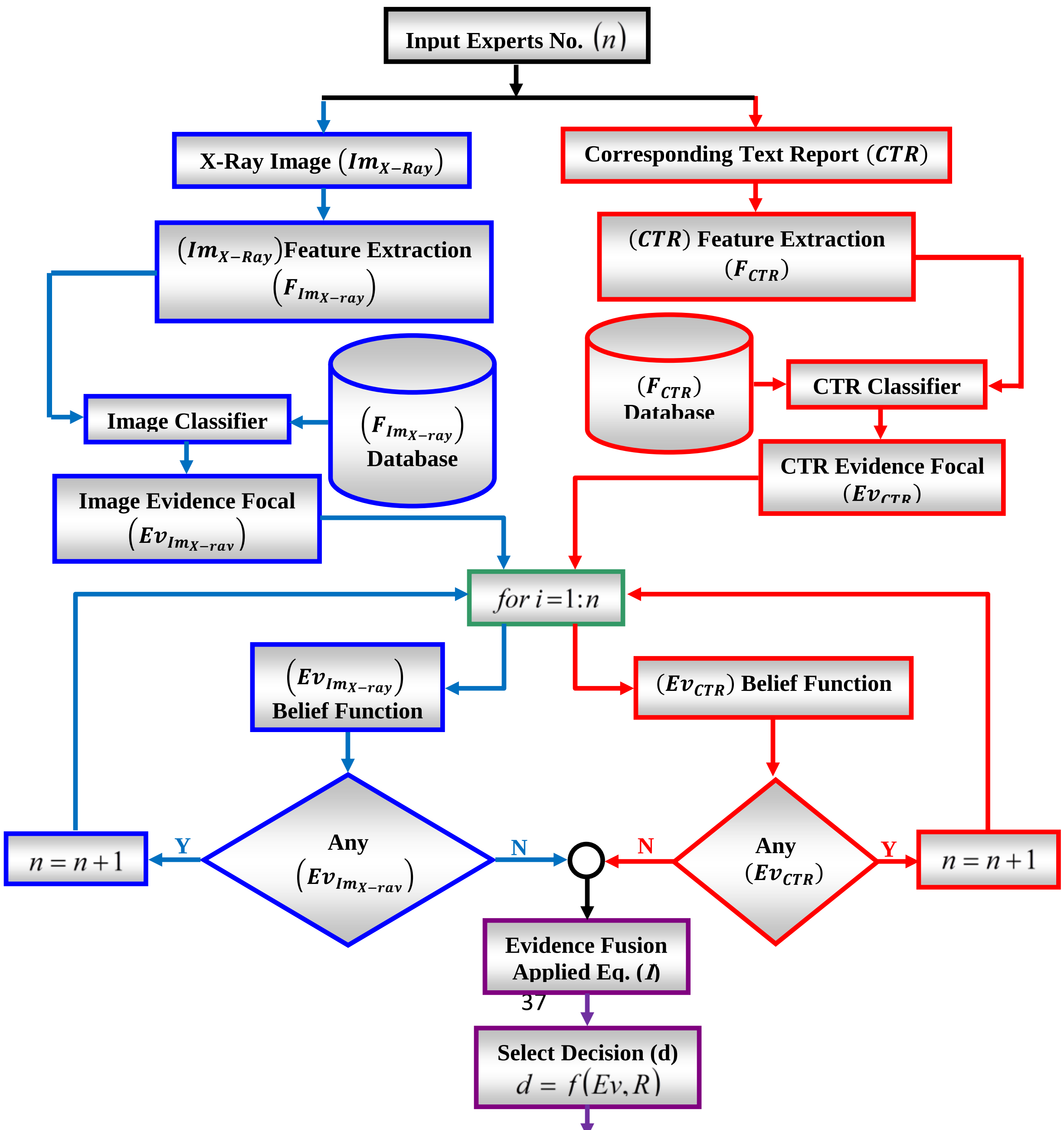


Fig. 2: Fusion of Evidence and Decision Making diagram

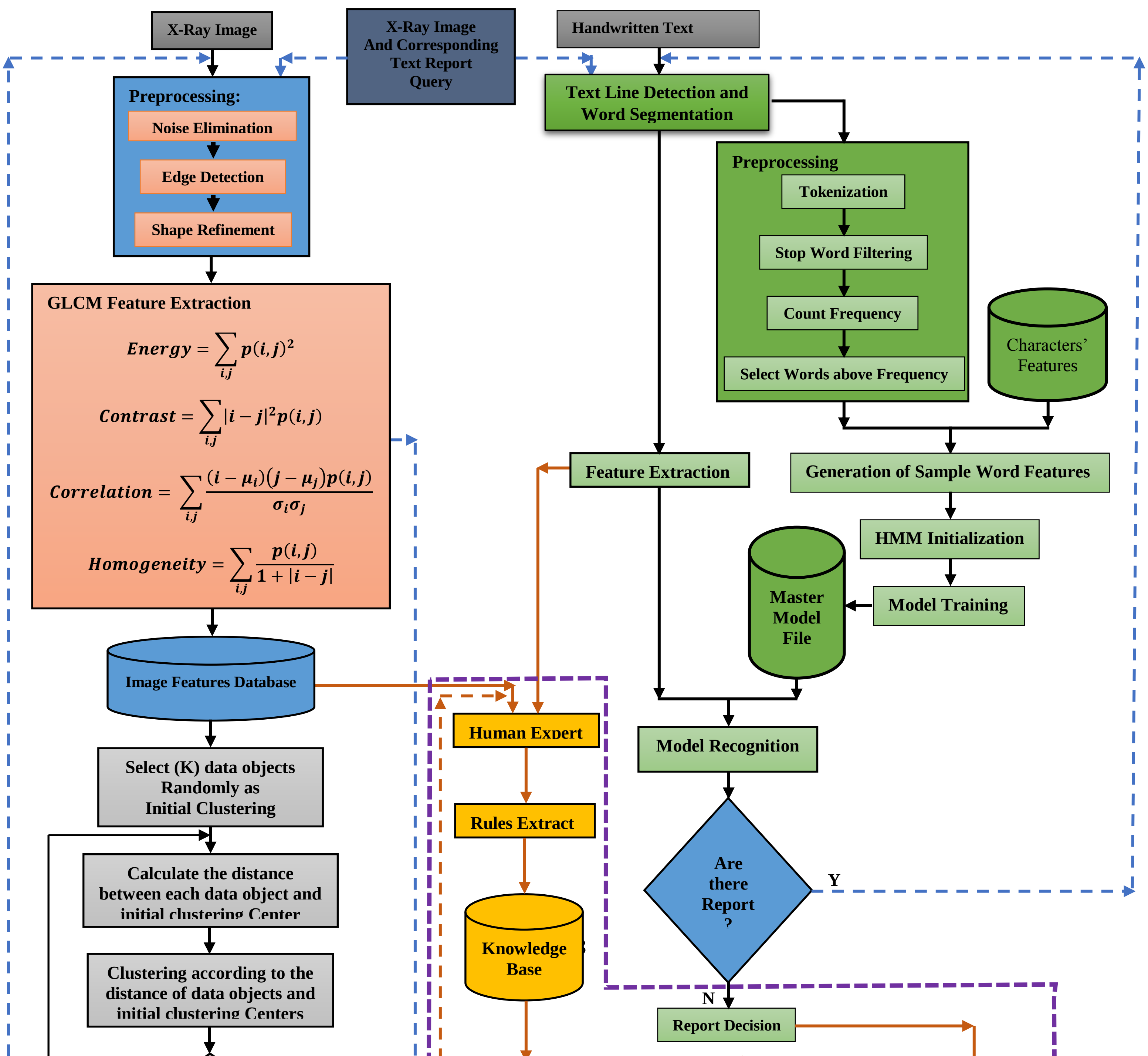


Fig. 3: Proposed method diagram.

Each normalized mass function $m_N(.)$ can represent by several function associated with belief function known as plausibility, commonality and disbelief. Plausibility function (pl) is the most important which represents the upper limits of uncertainty whereas the belief function (bel) represents the lower limits. Each of (pl) and (bel) functions are in one to one correspondence, they may be obtained from each other through linear transformation.

4. Experimental and Results:

Samples from X-Rays image with corresponding text reports had been collected as shown in figure 4.

Fig. 4: samples of X-ray Images and its corresponding reports.

Images and Corresponding Text Reports are fed to the proposed system, then both of them take its paths consecutively to extract its features. Images features are extracted and stored in database after image processing stage as shown in figures 5 and 6. Figure 7 illustrate assertion message for storing of image features successfully.

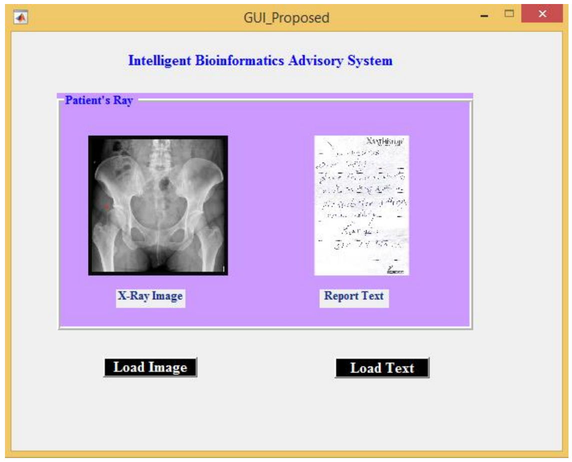


Fig. 5: main GUI for proposed system.

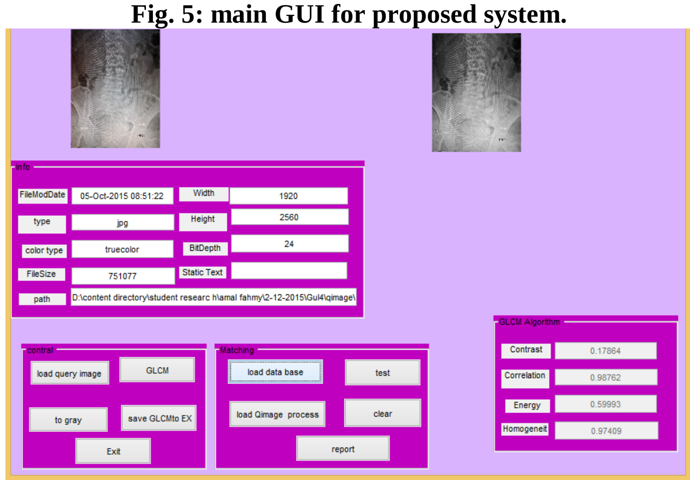
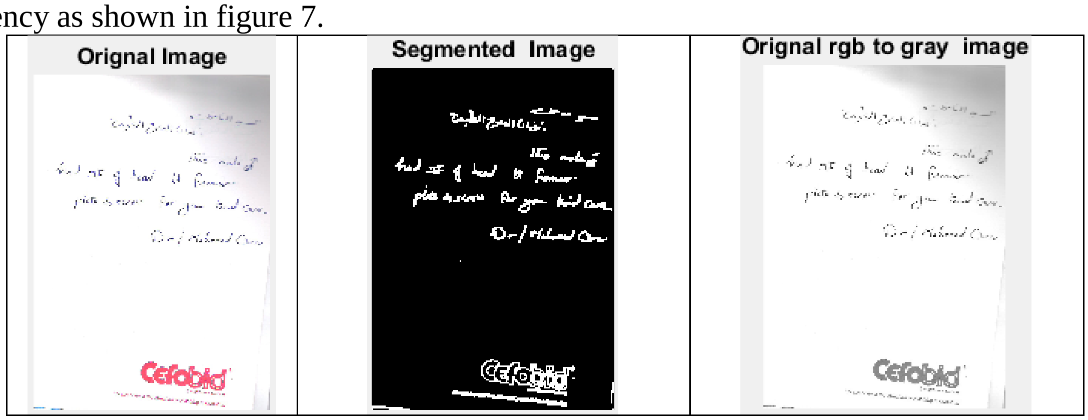


Fig. 6: X-ray image feature extraction

Image's features are classified into categories by using probabilistic evidence classification system which helps to retrieve the images in the fastest time.

The corresponding report for each X-ray image is addressed during the second stage namely text processing. In this stage, semi-continuous SHMMs that statistical modeling of handwriting is used to extract the features from corresponding report passing through several steps such as tokenization, stop responding word filtering, count frequency as shown in figure 7.



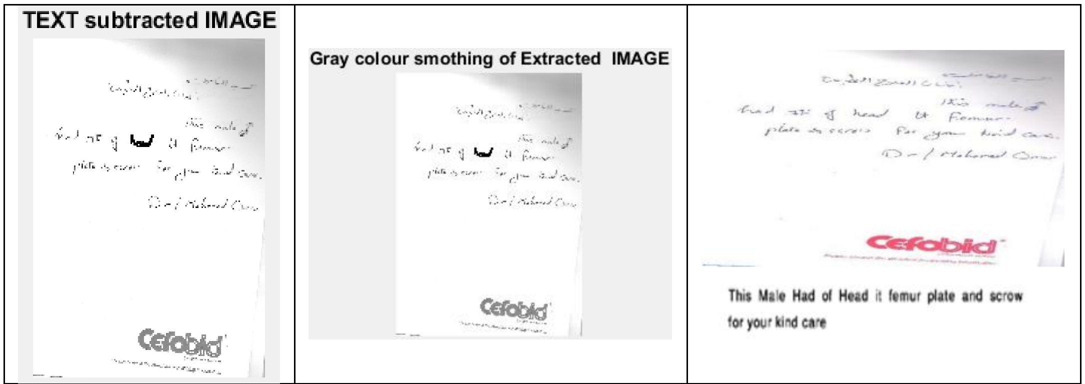


Fig. 7:SHMM handwriting report

The advisory part is achieved by placing the knowledge base rules. These rules are used both of expert men and the databases of image's features and automated text reports. When the query about treatment steps for Patients case based on X-ray images and the attached report, is fed to the proposed system and extract features of the image and convert the report to the computerized report, as previously explained. The rule is derived from both of the image and the report decisions as:

$$If\ (Im \in Cl_i) \cup \left(Im\ is\ labeled\ Im_j\right) \cup \left(Rep_{Text}\ is\ Rep_{Text_k}\right) Then\ Phy_{Th} is\ Rep_{Text_n}$$

The matching between the query rule and the knowledge base of the proposed advisory system determines the method of treatment as shown in Figure 8.

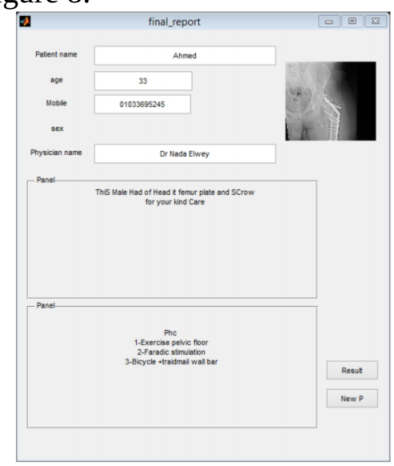


Fig. 8: final report

5. Conclusion

This paper presents a novelty advisory system in the medical field to making a decision based on image processing and text processing. In the image processing phase, the proposed system is fed by X-ray images, which used for extraction of their features that are classified according to their distinctive and stored in a database. In the second phase, text processing will be fed by the corresponding handwritten report which processed by hidden Markova model to convert it to automated report.

Then the results which stored in databases will be displayed on the expert that helped in the formation of the knowledge base. The results which were extracted from the proposed system are showed that it's convincing where, many of the patient's cases are applied and extract reports of treatment steps, which matched with the experts in the field.

References

[1] Barlett PF, editor. Agricultural decision making: Anthropological contributions to rural development. Academic Press; 2016 Jan 22.

[2] Abbott A. The system of professions: An essay on the division of expert labor. University of Chicago press; 2014 Feb 7.

[3] Jelassi T, Klein MR, Mayon-White WM, editors. Decision Support Systems: Experiences and Expectations: Proceedings of the IFIP TC8/WG 8.3 Working Conference on Decision Support Systems: Experiences and Expectations, Fontainebleau, France, 30 June-3 July 1992. Elsevier; 2014 Jun 28.
[4] Dennis A, Wixom BH, Tegarden D. Systems analysis and design: An object-oriented approach with UML. John wiley & sons; 2015 Mar 2.
[5] Fehr-Duda, Helga, and Thomas Epper. "Probability and risk: Foundations and economic implications of probability-dependent risk preferences." (2011).
[6] Herzog I. The potential and limits of optimal path analysis. Computational approaches to archaeological spaces. 2013 Aug 31; 186.
[7] Cobb LK, Anderson CA, Elliott P, Hu FB, Liu K, Neaton JD, Whelton PK, Woodward M, Appel LJ. Methodological issues in cohort studies that relate sodium intake to cardiovascular disease outcomes: a science advisory from the American Heart Association. Circulation. 2014 Mar 11; 129 (10):1173-86.
[8] McKeever S, Ye J, Coyle L, Dobson S. Using dempster-shafer theory of evidence for situation inference. InEuropean Conference on Smart Sensing and Context 2009 Sep 16 (pp. 149-162). Springer, Berlin, Heidelberg.
[9] Smets P, Kennes R. The transferable belief model. Artificial intelligence. 1994 Apr 1; 66(2):191-234.
[10] Dubois D, Prade H. Possibility theory: an approach to computerized processing of uncertainty. Springer Science & Business Media; 2012 Dec 6.
[11] Lin CH, Chen RT, Chan YK. A smart content-based image retrieval system based on color and texture feature. Image and Vision Computing. 2009 May 4; 27 (6):658-65.
[12] Liu Q, Sung AH, Chen Z, Xu J. Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images. Pattern Recognition. 2008 Jan 1; 41(1):56-66.
[13] Sahu SS, Panda G. A novel feature representation method based on Chou's pseudo amino acid composition for protein structural class prediction. Computational biology and chemistry. 2010 Dec 1; 34(5-6):320-7.
[14] Mark S. Nixon and Alberto S. Aguado. Feature Extraction and Image Processing. InAcademic Press 2008 (p. 88).
[15] Leydesdorff L, Wagner CS. International collaboration in science and the formation of a core group. Journal of informetrics. 2008 Oct 1; 2(4):317-25.
[16] Alexandratou E, Yova D, Gorpas D, Maragos P, Agrogiannis G, Kavantzas N. Texture analysis of tissues in Gleason grading of prostate cancer. InImaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues VI 2008 Feb 28 (Vol. 6859, p. 685904). International Society for Optics and Photonics.
[17] Khuzi AM, Besar R, Zaki WW. Texture features selection for masses detection in digital mammogram. In4th Kuala Lumpur International Conference on Biomedical Engineering 2008 2008 (pp. 629-632). Springer, Berlin, Heidelberg.
[18] Nithya R, Santhi B. Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer. International journal of computer applications. 2011 Aug 31; 28 (6):21-5.
[19] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics. 1973 Nov (6):610-21.
[20] Smets P, Kennes R. The transferable belief model. Artificial intelligence. 1994 Apr 1; 66(2):191-234.
[21] Wienecke M, Fink GA, Sagerer G. Toward automatic video-based whiteboard reading. International Journal of Document Analysis and Recognition (IJDAR). 2005 Jul 1; 7(2-3):188-200.
[22] Fink GA. Markov models for pattern recognition: from theory to applications. Springer Science & Business Media; 2014 Jan 14.
[23] Schwartz R, Chow YL, Kimball O, Roucos S, Krasner M, Makhoul J. Context-dependent modeling for acoustic-phonetic recognition of continuous speech. InICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing 1985 Apr 26 (Vol. 10, pp. 1205-1208). IEEE.

[24] Nithya R, Santhi B. Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer. International journal of computer applications. 2011 Aug 31; 28(6):21-5.
[25] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics. 1973 Nov (6):610-21.
[26] Yager, Ronald, Mario Fedrizzi, and Janusz Kacprzyk. "Advances in the Dempster-Shafer theory of evidence." (1994).
[27] Smets P, Kennes R. The transferable belief model. Artificial intelligence. 1994 Apr 1; 66(2):191-234.
[28] Denoeux T. A neural network classifier based on Dempster-Shafer theory. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans. 2000 Mar; 30(2):131-50.