Towards Enhanced Accuracy in Medical Diagnostics - a Technique Utilizing Statistical and Clinical Data Analysis in the Context of Ultrasound Images

Sweta Sneha, PhD
Kennesaw State University
ssneha@kennesaw.edu

Rajeev Agrawal, PhD
G.L. Bajaj Institute of Technology and Management
director@glbitm.org

Abstract

Biomedical imaging remains as a critical thrust area in the healthcare domain with colossal challenges and opportunities associated with analyzing and intelligent utilization of the enormous amounts of data. This research explores the domain of ultrasound images from the perspective of statistical and clinical data modeling and analytics. We propose a technique towards enhancing diagnostic accuracy of the current state and prediction of future states in the context of anomalous/cancerous tissue growth. The proposed technique utilizes both statistical and clinical data associated with ultrasound scans of the cancerous tissue in developing a probabilistic data model towards enhanced inferences with respect to current state of disease and future estimates of survivability. Preliminary evaluation of the proposed technique shows that combining statistical parametric estimates with clinical rule based prediction yields a more accurate model of disease diagnosis. One of the implications of the proposed technique is the potential of practicing evidence based medicine rooted in factual statistical and clinical data thereby reducing variability with respect to diagnostics and corresponding treatment.

1. Introduction

Analysis of ultrasound medical image is one of the widely used diagnostic methods for the assessment and classification of tissue growth as malignant or benign. [1]. Ultrasound echo signal are formed by summation of signals from randomly located sites through backscattering of radio waves inside the medium being imaged. The term radio wave in the current research is used in context of the backscattered echo received as an envelope. This represents the reflected electric field generally termed as ultrasonic backscattering [13, 14].

The ultrasound echo signals received from randomly located sites within the body are random in nature and are therefore modeled through probabilistic frameworks [2, 3]. Scattering and shadowing of radio waves inside the medium further add to the randomness of the received echo signal envelope [4].

Analysis of these random echo signals obtained via ultrasound scans provides a key data point towards making clinical predictions related to stage, grade, and thickness of anomalous tissue growth and corresponding survival of the patient. In recent years, there has been a shift from rule based purely clinical diagnosis to diagnostics utilizing statistical models for parametric analysis of ultrasound images.

The envelope of backscattered echo is used to extract the statistical parameters based on best fit models. In previous studies, several parameter combinations of complex probability distribution models have been used to capture the random statistics of these backscattered ultrasound echo signals. Rayleigh, Nakagami, and Rice are some of the widely used probability distributions that have been utilized for parameterizing the characteristics of ultrasound images [5, 6]. The traditional handcrafted rule-set with image segmentation approaches have been replaced with classifier techniques using Markov models for object detection and texture description [7, 8].

Building commercial biomedical imaging solutions based upon fully probabilistic models has become a major thrust area in medical imaging. It aims at understanding the effect of changes in clinical features of body tissues/cells on recovery/survival pattern of a representative group of patients. The probabilistic data models can be leveraged for drawing medical diagnostic inferences such as time to survival, shift from one stage to another, chances of recurrence etc. However, a critical challenge is the existing variability in assessment of ultrasound scans
and resultant diagnosis associated with statistical data modeling versus clinical rule based approach. Sole dependency on either classical rule based assessment or statistical parametric approach could lead to Type 1 and Type 2 errors in diagnosis and resultant treatment of the patients.

Given this context, the specific research question addressed by the current research is: “How can we enhance the accuracy of diagnostic inference and reduce variability in drawing conclusions based on assessment of ultrasound images of anomalous tissue growth?” We address this research question in the following three phases:

- **First phase** involves evaluation and selection of the best probability distribution function for parametric estimation of the ultrasound images. We utilize best curve fitting method to select Inverse Gaussian Distribution (IGD) as the model to capture the random nature of ultrasound echo signal envelope. IGD is leveraged to develop a statistical data model for analysis of ultrasound images. (Section 2)

- In the second phase the statistical data model based on IGD is utilized to obtain parametric estimates associated with multiple ultrasound images. Additionally, the historic diagnostic data associated with these ultrasound images are utilized in developing a mapping of observed clinical variables to the obtained parametric estimates. This information of estimated parametric values mapped to the clinical variables is leveraged in diagnostic evaluation of ultrasound images with respect to current/future states of the disease. The details of the cases used in deriving clinical inferences through statistical parametric estimation are presented. (Section 3)

- The third phase introduces the use of Markov State Model to trace the future condition of the patient. In the third phase the proposed technique integrating statistical and clinical approach is developed. Statistical information obtained in phase two is infused with clinical parameters through conditional probabilistic decision rules to derive more accurate inferences. The algorithm utilizing probabilistic rules and the corresponding analytical data model underlying the proposed technique are rooted in clinical and statistical characterization of ultrasound images. (Section 4)

- Preliminary evaluation of the proposed technique via ultrasound images taken over a nine week period shows that the proposed technique provides enhanced accuracy in diagnostic prediction and can potentially lead to reduction in subjective variability associated with sole reliance on classical clinical diagnosis approach. (Section 4)

The contributions of the current research are multifold. First, it proposes a technique that integrates the conventional clinical decision and statistical probabilistic modeling towards enhanced diagnostic accuracy in the context of ultrasound images. Second, it provides a preliminary validation of the potential of the proposed technique in increasing the accuracy of medical diagnostic in the context of ultrasound images. Third, it provides a model that is not only extensible for future studies but can also be applied to other types of biomedical imaging data if the data sets are predicting some probable estimates. Fourth, it provides a discussion of the practical implications of the current study. Fifth, it discusses the limitations of the current study and provides clear guidelines for future research (Section 5).

2. Best Fit Statistical Model for Parametric Estimation of Ultrasound Images

Ultrasound images obtained through 2D scan are normally represented in mode B where amplitude of the echo is directly related to the brightness of the dots. Average power of the echo signal envelope is random and contains overall statistical information about the scanned tissues. Ultrasound echo signal are formed by summation of signals from randomly located sites through backscattering of radio waves inside the medium being imaged. The signals which are reflected back from such medium are random in nature and can be modeled through different probability distributions and their statistical properties. These models are useful in relating medium parameters to the echo images and in characterizing the scattering medium.

Figure 1a presents a generic ultrasound image in B scan mode, followed by Figure 1b that presents an Image Density Plot of corresponding to the ultrasound image depicted in Figure 1a. Figure 1b shows the characterization of the statistical parameters associated with the ultrasound image in Figure 1a.

![Figure 1a: An Ultrasound Image in B Scan Mode](image-url)
Lognormal distribution (LD) has been shown to most closely characterize this random signal envelope. [9]. However, its complex and intractable expression led the researchers to explore other distributions like Weibull, Gamma, and other log-logistic distributions. The applicability of these distributions changes from case to case depending upon the presence of local or global variation of the scatters in the scanned area. A distribution with long range correlation can become a best fit model. Inverse Gaussian Distribution (IGD) is one such available solution [10] which provides more flexibility and wider coverage of range in terms of capability of representing echo envelope. The comparative plot of probability distribution function as depicted in Figure 2 (a, b, c, d) shows that IGD gives a closer approximation to log normally distributed echo signal envelope.

The probability distribution function of IGD with two characteristic parameters written as $\theta$ and $\lambda$ is presented in equation 1 [10] as:

$$\delta_y(y) = \sqrt{\frac{\lambda}{2\pi}} y^{-3/2} \left(\frac{\lambda(y-\theta)^2}{2\theta^2} y \right), \quad y \geq 0$$

The data model presented in equation 1 is leveraged for parametric estimation of ultrasound images. The presence of cancer as depicted via an unwanted growth in the ultrasound image, leads to an exponential curve of the echo envelope [15]. Since the proposed study focuses on analyzing/characterizing tissue growth in terms of malignant or benign, hence the growth curve is considered a key factor in the proposed model.

### 3. Clinical Inferences via Parametric Estimation of Ultrasound Images

Once we know the probabilistic distribution of the ultrasound echo signal envelope, a mapping is done between the characteristic statistical parameter set $(\lambda, \theta)$ with the clinical factors associated with the patients’ conditions. This mapping is based upon fitting lower and upper markings of the parameter values and relating them to the clinical variables such as: type of tissue (Malign/Benign), size of tumor.
(small, medium, large). This mapping of the statistical parameters to the clinical variables is leveraged in assessing and drawing future inferences from ultrasound images.

In order to utilize the statistical mapping, the ultrasound scans need a parametric estimation. Statistical parametric estimates associated with ultrasound images are obtained via curve fitting between distribution of the ultrasound images and standard distribution of IGD. The best fit point gives the estimated values of the characteristic parameters. These values associated with parametric estimations are used to draw clinical inferences such as: Type, Size, Stage, Grade, and Growth Rate of tumor through the mapping discussed earlier in this section.

Table 1 (below) presents five sample clinical inferences obtained through estimated parametric data analysis of ultrasound images. The parametric data associated with each of the five scans are mapped to clinical factors pertaining to Type, Size, Stage, Grade, and Growth Rate of Malignant Tumor.

Table 1: Mapping Statistically Estimated Parametric Values Associated with Ultrasound Scans to the Clinical Factors

<table>
<thead>
<tr>
<th>Scan</th>
<th>Estimated Value</th>
<th>Clinical Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \lambda = 0.59832 ) ( \theta = 5.07841 )</td>
<td>M - S - I - 1 - Linear</td>
</tr>
<tr>
<td>2</td>
<td>( \lambda = 0.37474 ) ( \theta = 20.0855 )</td>
<td>M - S - I - 1 - Linear</td>
</tr>
<tr>
<td>3</td>
<td>( \lambda = 1.58205 ) ( \theta = 2.71828 )</td>
<td>M - S - I - 1 - Linear</td>
</tr>
<tr>
<td>4</td>
<td>( \lambda = 12.9644 ) ( \theta = 9.7889 )</td>
<td>B - L - A - 2 - Unimodal</td>
</tr>
<tr>
<td>5</td>
<td>( \lambda = 29.4848 ) ( \theta = 8.3728 )</td>
<td>B - L - A - 2 - Unimodal</td>
</tr>
</tbody>
</table>

Type: Malign (M), Benign (B); Size: Small (S), Medium (M), Large (L); Stage: Initial (I), Advanced (A)

4. Technique Combining Statistical and Clinical Method for Enhanced Accuracy in Medical Diagnostic via Ultrasound Images

A conventional staging process, (which begins once a malignant cancerous growth is diagnosed) involves assessing patients’ current state, the extent of the cancer, and corresponding treatment options. The healthcare team in collaboration with the patient considers multiple factors during staging. Accurate assessment of the ultrasound image is one of the factors that provide key data points towards staging. The proposed technique provides an integrated (statistical and clinical) assessment of the data associated with the ultrasound images with the objective to improve medical diagnostic in the context of ultrasound images. The parametric values obtained via statistical inferences have minimum deviation error from a real data set [16, 17]. While the proposed model provides a richer data analysis in the context of ultrasound images via integrating clinical and statistical data, the existing models take into account either a clinical trend or a statistical approach [4, 6, 9, 11, 16, 17, 20].

The following paragraphs present an integrated approach that utilizes dependent conditional probability where statistical information is infused with clinical data obtained from medical experts for a more accurate diagnosis. Consider the case of growth of a malignant tumor that goes through four states. State N represents normal growth of the body tissues. Cancerous growth has been represented by three states viz. state A, state B and state C ultimately leading to death event (state D). Figure 3 shows a Markov Model with various probable state transitions and their corresponding transition probabilities. These probabilities are based on historical data pertaining to inferences made by medical practitioners for a given state. The objective of this Markov model is to convert rule based clinical inference into probabilistic values. These values are infused with probabilities obtained from parametric analysis of IGD density plot to obtain better accuracy in diagnosis and treatment.

Figure 3: Markov Model Showing States and Transitional Probabilities of a Malignant Growth
The transition probability matrix for single state transition based on the Markov model (as depicted in Figure 3) can be written as:

\[
\begin{bmatrix}
N & A & B & C & D \\
N & 5/6 & 1/6 & 0 & 0 & 0 \\
A & 1/3 & 1/3 & 1/3 & 0 & 0 \\
B & 0 & 1/4 & 1/2 & 1/4 & 0 \\
C & 0 & 0 & 1/3 & 1/3 & 1/3 \\
D & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\] (2)

**4.1. Integrated Diagnosis Technique**

In the previous section, we have presented statistically derived clinical diagnosis using IGD as a better fit model. However, the importance of rule based clinical diagnosis by medical practitioners can’t be ignored [11]. In the proposed technique, once the state of the patient is known through a combination of statistical and rule based decision making, the transition matrix (equation 2) can be used to obtain a conditional probability where the statistical probability is dependent upon clinical parameters.

The proposed integrated diagnostic technique is based on conditional probabilities obtained from replacing transition probabilities given in equation 2. Statistically computing a patient’s diagnosis is based on finding the parametric estimates associated with a given case at a particular instant in time. The parametric values are obtained from density plot of the scans taken at different intervals of time. Curve fitting of best/closest model with density plot of the ultrasound scan is used for parameter estimation as discussed in section 2 and 3. In this research we have utilized IGD as the best fit model.

Accurate inferences obtained from analysis of ultrasound images provides a critical data point that is utilized in making clinical diagnosis pertaining to treatment plan and forward/backward progress of the tissue growth in terms of stage, grade, size etc. Each of the probable stages and possibility of transition to other stage is assigned probabilities as per the transitions shown in figure 3. These probabilities are calculated by comparing the estimated values of the parameters with the corner values of the range assigned to each stage. An estimated value lying in exactly middle of the range is given a probability of 1 that the patient remains in the same stage and equal transition probability of 0.5 to both adjacent stages (forward and backward). A value closer to lower bound has reduced probability of remaining in the same state and higher probability of transition to lower stage. Similar is the case with values closer to upper bound.

However, the prediction based upon this statistical technique may not always match with the rule based inferences made by the medical practitioners and it is difficult to select one of them in case of mismatch. We therefore propose to enhance the diagnostic prediction by integrating both these approaches through a conditional probability based technique. The condition can be applied on the clinical information in terms of statistically calculated probability or vice versa.

Let \( P_e \) be the probability assigned through clinical diagnosis/inference by the medical practitioner and \( P_s \) be the probability obtained through statistical analysis of the scan. Integrated observation is represented as ‘O’. For the probability \( (p_t) \) at a particular instant of time, the observation is written as \( O_t \). Taking a particular case ‘C’ and observing at a particular instant of time the observation is recorded as \( O_t(C) \). Then the final integrated observation can be written as:

\[
O_t(C) = P(p_t = P_e | P_c) \text{ OR } P(p_t = P_c | P_e)
\] (3)

Applying Baye’s rule [12], equation 3 is isolated as:

\[
O_t(C) = P(p_t = P_e, P_s) / P_c \text{ OR } O_t(C) = P(p_t = P_c, P_s) / P_e
\] (4)

Using time separation on a particular case (patient), we can write:

\[
\delta_t(C) = P(P_s, t_1, t_2, t_3, ... | P_t = P_e) \text{ OR } \sigma_t(C) = P(P_s, t_1, t_2, t_3, ... | P_t = P_c)
\] (5)

The observation can now be calculated as:

\[
Q(C) = \frac{\delta_t(C)\sigma_t(C)}{\sum_{i=0}^{N} p_i(P_t) \text{ OR } p_i(P_e)}
\] (6)

These observation values are compared with both \( P_e \) and \( P_s \) and a simple rule of closeness of minimum deviation is followed to take the final decision regarding stage of disease and probability. For instance, if the clinical inferences indicate a diagnosis with certain probability \( a \), a new probability based on statistical data analysis is obtained by taking into account the clinical probability as a conditional probability. Since we have used the conditional probability which is dependent, the confounding issue of data independence is mitigated.
4.1.1. Preliminary Assessment of the Proposed Integrated Diagnostic Technique

In order to validate the utility of the proposed technique in enhancing diagnostic accuracy while reducing discrepancy between purely clinical rule based and purely statistical diagnostic approach, a preliminary evaluation is performed by considering the case of a patient suffering from malignant tumor. Four scans at a regular interval of three weeks were taken and analyzed for diagnosis using classical rule based approach and statistical analysis.

The scans were obtained from a local hospital of a male patient, age between 45-47 years. The patient was admitted with a complaint of pain in lower abdomen. Ultrasound imaging of the lower abdomen focusing on kidney and lever of the patient was taken. Based on the assessment of the ultrasound by a healthcare team the patient was diagnosed with malignant tissue growth. The results of the treatment are observed for the same patient at same clinic but at different time intervals. The healthcare team consisted of a radiologist, pathologist, and an expert physician.

The main objective of the preliminary assessment is to establish the feasibility of the proposed approach in enhancing the accuracy of diagnostics via ultrasound images. Although the initial validation of the proposed approach considers a relatively small sample size, (4 scans over nine weeks), future work will enhance the robustness of the validation by considering a larger sample size consisting of multiple cases and diseases. Table 2 presents clinical and statistical inferences drawn over a nine week period along with the corresponding probabilities (P_s) and (P_c) in the context of test cases.

The clinical and statistical values in table 2 show the discrepancy in diagnosis via purely statistical and clinical approaches for two time periods, i.e., t1 and t3, in terms of stage of disease. Hence we apply the proposed integrated diagnosis technique to remove this discrepancy and enhance diagnostic accuracy. The set of probabilistic data values for P_c and P_s as shown in table 2 are presented as:

\[ P_s = \{0.35, 0.7, 0.4, 0.75\} \]
\[ P_c = \{0.5, 0.5, 0.3, 0.3\} \]

For integrated diagnosis, consider the case of \( P(P_s|P_c) \) i.e. statistical diagnosis subject to the condition of known clinical diagnosis. Both the probabilities are assumed to be independent. Using integrated diagnosis algorithm, the observation set \( O_t(C) \) is obtained as:

\[ O_t(C) = \{0.218, 0.437, 0.25, 0.468\} \]

On comparing with corresponding values for \( P_c \) and \( P_s \), the diagnosis through \( P_s \) (stage A) at time t1 and that through \( P_c \) (stage A) at t3 is taken as final decision. For the scans taken at time t2 and t4 the diagnosed decisions are same through both the techniques. Only the probability closest to observation value selected. The final data set associated with diagnosis in terms of stage and probability is given in table 3 which establishes the utility of the proposed technique in eliminating discrepancy in medical diagnosis.

Table 3: Diagnosis via Integrated Technique

<table>
<thead>
<tr>
<th>Stage</th>
<th>t1 initial</th>
<th>t2 3 weeks</th>
<th>t3 6 weeks</th>
<th>t4 9 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_s )</td>
<td>Stage A</td>
<td>Stage B</td>
<td>Stage B</td>
<td>Stage A</td>
</tr>
<tr>
<td>( P_c )</td>
<td>0.35</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

5. Conclusion

Ultrasound medical image is one of the widely used diagnostic tools used in the assessment and classification of a tissue growth as normal or abnormal. Ultrasound echo signals are formed by summation of signals from randomly located sites through backscattering of radio waves inside the medium being imaged. Analysis of these random echo signals obtained via ultrasound scans is utilized in making clinical predictions related to stage, grade, and thickness of anomalous tissue growth and corresponding survival of the patient.

Medical diagnosis regarding patients’ conditions in terms of stage of growth, survivability estimation, and corresponding treatment options via rule based
clinical assessment is a complex process spanning multiple factors and stakeholders. The potential subjectivity in this approach can lead to individual variations in diagnosis and treatment. In recent past, there has been a shift towards statistical analysis of random signals present in medical images for drawing clinical inferences and to predict the chances of survival. Both the rule based and statistical techniques are based on independent data analysis and despite of the inherent advantages, can sometimes lead to discrepant diagnostic decisions. Hence, sole reliance on any one of them can potentially lead to erroneous diagnosis leading to incorrect treatment plan.

The current research investigates this critical issue with the objective to develop a technique towards enhancing the accuracy of medical diagnosis via ultrasound images and reducing discrepancy in the subjective diagnosis associated with purely rule based and purely statistical approaches. With that goal in mind, we present a technique which infuses the rule based inferences drawn by medical practitioners with statistical probabilistic analysis in the context of tissue characterization via data analysis of ultrasound images. In addition, the approach of integrating clinical and statistical data is widely reported in the literature to provide new insight, features, and information, in the context of inferring clinical predictions.

The underlying algorithm for the proposed technique utilizes conditional probabilistic data models. Statistical diagnosis is obtained by the mapping of known clinical variables to the parametric estimates obtained by curve fitting of image density plots of ultrasound images via IGD. Since the model takes the estimation from the best fit/closet model (IGD), the parametric values received in case of statistical inferences have minimum deviation error from a real data set [16, 17]. Whereas the proposed model, can lead to improved clinical diagnostic via a richer analysis based on integrating clinical and statistical data, the existing models take into account either a clinical trend or a statistical approach [4, 6, 9, 11, 18, 19, 20]. Additionally, we utilize Markov model to represent the transition between different stages of the disease along with estimated transition probabilities.

We performed a preliminary validation of the proposed integrated diagnosis technique by considering the case of a malignant tumor observed over nine weeks. Four ultrasound scans at different time intervals over nine weeks were used for evaluation of the proposed technique. The results establish the utility of the proposed scheme in improving diagnostic accuracy and reducing the discrepancy between rule based and statistical diagnosis in the context of tissue characterization and corresponding diagnosis of a patient.

5.1. Implications of Current Research and Directions for Future Research

One of the practical implications of the current research is the potential for promoting evidence based medicine in diagnosis and corresponding treatments. Since the proposed technique is rooted in conditional probabilistic model drawn from factual data associated with statistical and clinical characteristics of ultrasound images, hence leveraging this technique by medical practitioners will possibly enhance accountable healthcare delivery and practice.

The current exploratory work provides preliminary validation of the feasibility of the proposed technique in enhancing accuracy in medical diagnosis. Further research is warranted to address the limitations of the current research and to assess the robustness of the proposed technique over a larger sample size. Few avenues for future research include but are not limited to the following:

- Due to the exploratory nature of the current research, the predictive model is kept simplistic. Hence we have focused on the clinical parameter associated with the stage of growth of the malignant tissue in the probabilistic data model. Future research is called on to extend the proposed probabilistic data model by incorporating other clinical parameters such as: grade, growth rate, tumor size, and patients' background information such as: age, ethnicity, medical history etc. in assessing the impact on diagnostic accuracy.

- This research can be used as a foundation for investigating the prospects of machine learning in medical diagnosis via ultrasound images. A data model based on statistical parametric estimates mapped to known clinical inferences with respect to ultrasound images has the potential to yield a machine learning environment for predicting medical diagnosis.

- Preliminary evaluation of the proposed technique (via a smaller sample size of 4 scans over nine weeks) validates its potential in enhancing diagnostic accuracy and reducing discrepancy in sole statistical and clinical based approaches. However, future research is necessary for thorough evaluation of the proposed technique.
and to test its robustness via a larger sample size consisting of multiple cases before it can be put to practice.

- The current research calls to investigate the development of visualization and decision support tools that utilize the proposed conditional probabilistic model for assisting in clinical decision making.

- Future research should also investigate an in-depth analysis of the comparative utility of the proposed integrated approach in enhancing clinical diagnostic in the context of ultrasound images with other existing approaches that utilize the assessment of ultrasound images in aiding clinical decision making.

- Although, the proposed model can theoretically be applied to other biomedical imaging data for analytics, as long as there are two data sets that are predicting some probable information. However, future research is required to assess the practical applicability of the proposed technique to other biomedical imaging data in addition to ultrasound scans.

Biomedical imaging remains as a critical thrust area in healthcare domain with colossal challenges and opportunities associated with analyzing and intelligent utilization of the enormous amounts of data. The investigation process has only just begun and our hope is that the current research will add value to this nascent domain and open doors for multiple avenues for research and collaboration potential across multiple disciplines such as information systems, engineering, biology, and medicine.

6. References


