International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Regional Language Support for Patient-inclusive Decision Making in Breast Cancer Pathology Domain

G. Johanna Johnsi Rani, Gladis D., Joy John Mammen

Abstract*: A Clinical Decision Support system (CDSS) is an application that analyzes data to help healthcare providers to make decisions and improve patient care. Clinicians use the CDSS to perform their routine tasks with computer-assistance. In the past, decision-making using CDSS was primarily oriented towards Clinicians but in recent times, shared decision-making with the patient is advocated. Shared decision-making focuses on encouraging patients to become informed and involved about their health-concerns and make right choices in discussion with expert Clinicians. In India, Breast cancer is the number one killer disease among women. The fast-growing breast cancer patient population demands development of a CDSS for the domain with patient-inclusive features. Medical documents generated in English by Medical practitioners may be understood only by patients with adequate medical knowledge and proficiency in English language. To benefit the regional-language patient population, a CDSS was developed with patient-inclusive features such as Risk assessment questionnaires and Pathology reports presented in Tamil to benefit the regional language-literate patients in the state of Tamilnadu. Translation resources for the domain such as Lexicon and Bilingual Dictionary are generated and used in Machine Translation (MT) of the reports in the CDSS. Translation of Pathology reports is performed by applying Natural Language Processing methods and Phrase-based translation approach and is refined using Synsets. The machine-translation by the CDSS was evaluated by comparing the CDSS output with output from a translation tool Anuvadaksh developed by Department of Information Technology, Government of India, and Google Translate. The outputs were also scrutinized by regional language experts and medical experts. The developed CDSS prototype is a pioneering effort to compile medical language resources for breast cancer pathology domain, and to present details to the patient in a language familiar to her. The regional language support would improve co-operation between the Clinician and patients for shared decision-making and enhance understanding in patients who would otherwise be passive due to the English language barrier. The CDSS with regional language could be used in hospitals in Tamilnadu and the implementation could be extended to other regional languages of India in the future.

***Keywords*: Clinical Decision support system, Breast cancer, Machine translation, Natural Language processing, Shared decision-making**

I. INTRODUCTION

Health-care must be patient-centric rather than Physician-centric and shared decision making is a key component of patient-centered health care. The need for shared decision-making can be better understood in the words of Goldbach paraphrased as follows: The patient best knows about herself and the details about her are needed by the Clinician, and the Clinician is the expert on the disease, from whom the patient needs information. Together, they can find the best approach to a clinical problem. [1] In India, shared decision making is not the norm. Considering research evidence and patient inputs for a holistic approach to decision-making is still not widely prevalent. High quality randomized controlled trials, observational studies, considering clinical expertise and requirements of patients would take time to implement in India. Till then, shared decision-making would remain a concept restricted to the academic discussions instead of being applied in practical health care. [2] Patients can be either be *passive, co-operative or active* and Clinical decision-making models in the light of patient participation can be categorized as *paternalistic model*, in which patients play a passive role, *professional-as-agent model*, in which the doctor makes the decision for the patient, *informed decision-making model* in which necessary information is shared with the patient and he makes the decision at the end and *shared decision-making model* in which the patient and the doctor are responsible for the final decision. [3]

In the Indian health care scenario, the concept of shared decision-making has to be considered in the cultural and literacy aspects. A qualitative study and ethical analysis on Healthcare Decision Making in Rural India, a decade ago brought to light, the glaring differences between shared decision making in the Western world and India. The Western culture emphasizes on autonomy of individuals and freedom of choice by the individual. But in India, autonomy is less individualistic and decision-making depends on the patient's family and sometimes on their community. Further, in India Doctors are considered as supreme intellectual beings with high social status, and patients express shyness, fear and hesitation in taking an equal role with the Doctor in decision-making. [4]

Manuscript published on 30 September 2019
* Correspondence Author
G. Johanna Johnsi Rani*, Department of Computer Science, Madras Christian College (Autonomous), University of Madras, Chennai, India. Email: johanna.g@mcc.edu.in
Gladis D, Principal, Bharathi Women's College (Autonomous), University of Madras, Chennai, India, Email: gladischristopher@gmail.com
Joy John Mammen, Professor and Head, Department of Transfusion Medicine & Immunohaematology, Christian Medical College, Vellore, India, Email: : joymammen@cmcvellore.ac.in

Another major constraint to shared decision-making by the patients in India, is their incapability to read and understand medical information due to illiteracy or lack of proficiency in English language. Patients cannot effectively participate in shared decision-making if they do not understand their health condition, the treatment options, and the possible side effects of each treatment option. [1] The cultural and literacy barriers can be addressed by raising the awareness in patients regarding the benefits of shared decision-making in terms of treatment and in boosting their confidence and wellbeing before and after treatment. The other possible measure is to provide regional language services to the patients, since majority of the patient population are proficient in their local languages. This would also enable the family members and caregivers of the patients to understand their health condition and support them in making treatment decisions and provide care to them in the future. Development and use of Clinical Decision Support Systems (CDSSs) with regional language support would improve patient-inclusiveness, patient understanding and co-operation with the Doctor during treatment. Earlier studies have shown that if patients with lower literacy are provided with health literacy information they participate well in decision-making. [5] In principle and in practice, patient-inclusive decision-making has several benefits of improved health and well-being, care and quality, and fulfillment of Legal duty towards patients. [6]

Majority of the hospitals in India have limited electronic resources and software tools for decision support and CDSS are used only in high-end hospitals. Medical practitioners generate hand-written or electronic reports in English, which cannot be understood by majority of the patient population. Tamil is an ancient, regional language of the natives of Tamil Nadu. Tamil language has a rich vocabulary but research efforts have been minimal until now with respect to building language resources in Medicine and translation of Medical documents from English to Tamil. To encourage and widen the usage of CDSS, and advocate patient-inclusive decision-making, BEACON (BrEAst Cancer in wOmeN) CDSS was developed with regional language support resources in Tamil, translation services and regional language interfaces as important components. The system uses custom-generated, domain-specific Lexicon and bilingual dictionary, translates pathology reports from English to Tamil and provides Risk assessment templates (General and Family History) to the patients along with Lifestyle recommendations to them in Tamil. These features aim at enhancing the patient's understanding of their health conditions thus encouraging their participation in shared decision-making.

The remaining content of this paper is organized as follows. Section 2 explains related works in CDSS and Shared decision-making, Section 3 gives the material and methods used in incorporating patient-inclusive features in the CDSS, Section 4 explains the results and evaluation of the work, Section 5 presents the conclusions and scope for future work.

II. RELATED WORKS

Clinical decision support systems are software tools that support a Clinician in decision-making and can be categorized into passive systems in which he can make an explicit request for support, semi-active systems that are invoked automatically to present information to the Clinician at his request or active systems that make decisions and interact without input or request by the user. They can also be categorized as simple systems that raise user alerts to midlevel systems that are prognostic calculators and complex systems that use Artificial Intelligence, data mining and statistical methods. CDSS that focus on Clinicians in breast cancer domain are those that provide support for breast cancer diagnosis [7, 8], screening [9, 10, 11], raise user Alerts at patient's risk level, Symptom checking and risk stratification [21 mac], treatment recommendations [12], and use a conjunction of clinical and pathological data with Genomic tools for treatment decisions [13]. Clinician-centric decision-making is centered around Medical aspects while Patient-centric CDSS must provide templates and reports that patients can understand and participate in decision-making.

The factors that influence Patient Participation in Shared Decision-Making are categorized into Patient-Related Factors (Literacy, patient's understanding of the disease), Disease-Related Factors (treatment options and outcomes), Factors Related to CDSS (User-understandable interfaces, templates and reports), Factors Associated with Health Care Settings (use of EHRs and CDSS) and the Factors Related to Health Provider Tasks. (Screening, Diagnosis, recommendations and references) [14]

Shared decision making (SDM) is the key to patient-centric care, because both the patient and the Clinician participate in medical decision making [4]. For a medical condition, there are several treatment options and more than one decision can be made. In a shared decision-making environment, patients can share their preferences, and choose their options about treatment, resulting in co-operation between the Clinician and the patient. [13]. SDM makes the pathologist-patient communication effective and eliminates possible misunderstanding between the two and the ignorance in the patient regarding the disease and treatment [6].

Treatment decisions are focused on patient welfare; hence patient involvement is of vital importance.. SDM tools / CDSS with SDM features must inform patients with regard to the risks, benefits, and trade-offs associated with a decision. Use of such tools reduce conflicts in decisions, increase the ease of decision making and revision of previous decisions [5]. Studies show that patients who are informed about their disease through decision aids would willingly participate in the decision-making process and make right decisions. [5]. SDM respects an individual's right to choose, yet it is clinically restrained by a well-informed and well-meaning Clinician who is committed to serve the patients. CDSS also provides additional benefits over traditional methods of decision-making, as they have the potential for individualized content, a greater degree of interaction, and scalability [14]. Machine Translation systems have been developed for Indian Languages since 1995. The translation systems for translation from English to one of the Indian languages are Anglabharti I (1991), II (2004) English to Indian languages (primarily Hindi) a general-purpose machine translation system with provision for domain customization,

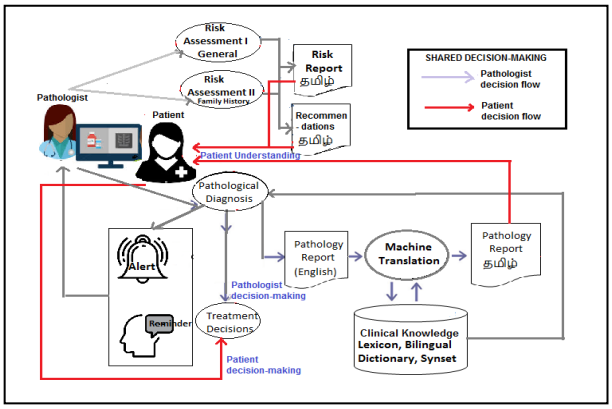
English to Indian languages, *Anusaaraka* (1995) from English to Hindi, *MaTra* (2004) converts complex English sentences into simpler sentences and translates the content into Hindi, *Mantra* (1999), English to Indian languages and Reverse developed for use in the Rajya Sabha Secretariat, *UCSG-based MT* English-Kannada, *UNL-based* (2003) Between English, Hindi, and Marathi, Tamil-Hindi, *Anusaaraka MT*, a Human Aided Machine Translation System English-Tamil HAMT, *Shiva* (2004) English to Hindi, *Shakti* (2004) English to Hindi, Marathi and Telugu, *Anuvaadak* English-Hindi, *English-Hindi Statistical MT*, *English-Hindi MAT*, *Hybrid MT system* English to Bengali, *Hinglish MT* system (2004) Hindi – English, English to Indian and Kannada to Tamil language-pair EBMT system (2006), *OMTrans* English-Oriya, English-Hindi EBMT system, *Machine Aided Translation* English to Hindi, *English-Tamil MT*, *Pictorial knowledge Based MT* (2010) *English to Assamese*, *English-Malayalam Statistical MT (*2009), *enTel* (2011) English to Telugu, *English to Indian Languages MT* (2007), Incorporating Syntactic and Morphological based MT English-Hindi, English-Telugu, *T2T MT* and Telugu-Tamil MT (2004), English-Urdu MT via Hindi (2009), *Vaasaanubaada* (2002), *Phrase based English - Tamil MT* (2011), Sentence Simplification System for English to Tamil (2011), Manipuri-English Bidirectional MT (2010) Manipuri-English and -English-Manipuri, English to Dravidian Language MT (2010) English to Malayalam, Anuvadaksh English to six other Indian languages i.e. Hindi, Urdu, Oriya, Bangla, Marathi, Tamil, and *Google Translate*. *AnuBharti* (2004) *Anubaad* (2004) and *Sampark* (2009). *Anuvadaksh* system developed in the English to Indian Language Machine Translation (EILMT) project, funded by the TDIL program of the DeiTY, Govt. of India translates English text into 8 Indian languages namely Hindi, Bengali, Marathi, Urdu, Tamil, Oriya, Gujarati and Bodo) [15].

Considering domain-specific translation systems, Thenmozhi D and Aravindan C proposed a Statistical machine translation system that translates textual content in the Agriculture domain from Tamil to English. [16] Dr. K. Narayana Murthy developed English-Kannada MAT system based on a transfer-based approach which is applied to the documents related to government circulars. Anuvaadak 5.0 system was developed for a general purpose of automatic translation from English-Hindi by Super Infosoft private limited, Delhi for specific domains like official, formal, agriculture, linguistics, technical and administrative using inbuilt dictionaries. The review indicates that there are no medical domain-specific translation systems.

With respect to translation approaches, Poornima C, Dhanalakshmi V, et. al proposed a preprocessing tool that converts complex sentences to simple sentences in English, using rule-based technique before translation to the target language Tamil [17]. R. Harshwardhan et.al proposed a framework for Phrase-based translation from English to Tamil using Translation memory and concept labeling [18]. Dhanesh N proposed a Conceptual Framework for Automated English to Tamil rule-based machine translation system [19]. S. Saraswathi et.al proposed a Bilingual translation system for English and Tamil applying Hybrid approach that uses Rule based Machine Translation and Knowledge based Machine Translation [20]. The developed CDSS is breast cancer pathology specific that translates Pathology report from English to Tamil using Phrase-based translation approach.

III. MATERIAL AND METHODS

Stephanie Medlock et. Al. proposed a two-stream model of information flow within clinical decision-support systems: Clinical stream that involves reasoning about the patient and the cognitive-behavioral stream that involves reasoning about the user. [22]. The first stream focuses on the doctor / CDSS that collects patient data, and interprets it using clinical knowledge for making a decision about the patient i.e. *Pathologist decision*. The second stream focuses on the presentation of recommendations to the patient to act on the decisions i.e. *Patient's decision* on Pathologist recommendations. In the Indian context, we consider an intermediary phase to bridge the gap between the Clinician and the patient's decisions to Patient's *understanding* by providing services that would break the language barriers between the two. When patient's understanding increases, cooperation in decision-making also increases.



Shared Decision-making in Breast cancer domain

The decision-flow for the domain presented in Fig. 1 indicates a set of medical treatment processes performed by the Pathologist with the help of CDSS. His expertise is applied to Pathological diagnosis, assessment of risks in a patient's health conditions and providing suitable recommendations regarding further treatment. The first aspect of patient's involvement is in answering the risk questionnaire in the CDSS, since it is in regional language that he understands. If Pathological diagnosis is recommended and performed, the Pathology report of the patient is generated in English, which is translated to Tamil and presented to enhance patient understanding.

A. Templates for User Interactions

A patient's involvement in decision-making is encouraged as the BEACON CDSS provides templates in regional language Tamil, one for risk assessment based on General parameters and another based on Family History. The Family history questionnaire is shown in Fig. 2 and Fig. 3. Both the Questionnaires can be answered by the patient guided by the Pathologist in case of doubt, resulting in shared assessment process too.

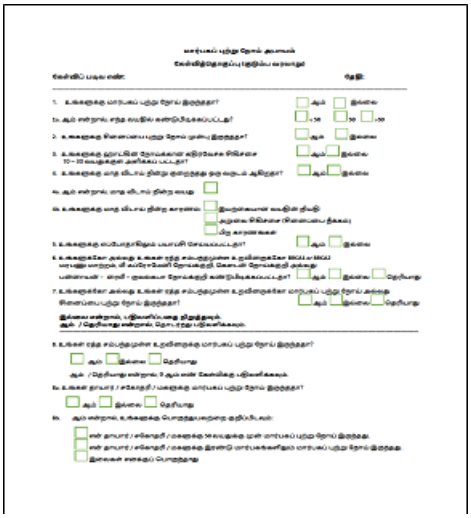


Fig. 1.Risk Assessement based on Family History –Pg.1

Fig. 2.Risk Assessement on Family History – Pg. 2

B. Building the Lexicon and Bilingual Dictionary

An important patient-inclusive feature in BEACON CDSS is the machine translation of Pathology report generated English by the Pathologist, to Tamil. Currently existing Medical resources in regional languages are the "English-Tamil Glossary of Medicine" by Tamil Virtual University, "Medical dictionary" translated from English to Tamil and Telugu compiled by D. Rambabu, and translated along with V.V. Rathnasree. and "தட்சநாயனார் தொகுத்து அளித்த மருத்துவ அட்டவணை", a compilation of Ayurveda and Siddha medicine in Tamil. [23, 24, 25, 26]. The resources in Tamil are not adequate to translate reports in breast cancer pathology domain. Hence a Lexicon of breast cancer terms and Bilingual dictionary of English and Tamil terms for each item in the Lexicon are custom-built. This was done with medical and language precision in mind and the resulting resources, are subsequently used in the machine translation process.

The Lexicon of Breast Cancer Pathology terms is generated using three approaches. The first approach applies Natural Language Processing steps namely Section segmentation, Sentence splitting and tokenization of the document contents to the dataset namely the Breast Cancer Pathology reports and the American Joint Committee on Cancer (AJCC) protocol and compiles unigram Medical Terms. The second approach is a set of Breast cancer pathology Glossary terms collected from online resources. In the third approach, the system allows the user to select Medical phrases from the dataset through online selection. The Lexicon generated consists of approximately 3,517 terms belonging to the domain and serves as the source to build the bilingual dictionary.

Tamil Dictionary and Thesauri of Medical terms for Breast Cancer Pathology are not available. The Tamil equivalent of the terms in the Lexicon were hence translated with the help of Tamil language experts, using Google Translate and a few other online Tamil dictionaries [27, 28] to build the bilingual dictionary. The terms in the Lexicon fall into the following categories - Translated words (unigram), Translated phrases (bigrams or more), Transliterated items and Non-translatable items and Abbreviations. An example of translated unigram is Tumour / Tumor, which is translated to கட்டி. Adjacent breast tissue is a phrase that is translated to அருகில் உள்ள மார்பக திசு. Medical terms that are not translatable without loss of meaning are transliterated. For example, Progesterone is transliterated to புரோஜெஸ்ட்டிரோன். A small subset of Standard Medical terms are non-translatable. For example, Pathological classification pTNM which constitutes of pT, pN, pM and cancer stage representations such as Tis(DCIS), T1a, etc. are not translated in order to retain their Medical significance in the text. Abbreviations are also not translated. The system provides an editor for the Language experts in Tamil and the Medical domain experts to check the preciseness of the medical terms in the target language and set the Gold standard for translation purposes. We have limited the translation to nearly 2393 important terms. In the current Lexicon, 891 are translated terms, 46 are Non-translatable terms, 36 are Abbreviations and 151 are Transliterated terms.

C. The Dataset

Machine translation was performed on 150 Breast Cancer Pathology reports obtained from a renowned hospital in Tamil Nadu, South India. Each report has five sections namely Specimen, Clinical, Gross, Micro and Impression as shown in a sample report in Fig. 4. The entire report content is translated by the CDSS.

1) 18462/12
SPECIMEN : Right MRM
CLINICAL : Carcinoma right breast.
GROSS : Specimen of right modified radical mastectomy measuring 19x11x7cm with nipple bearing skin measuring 14x9cm and attached axillary pad of fat measuring 5.5x4.1cm. The nipple and areola appears to be uninvolved. Sectioning reveals a tumour measuring 4x3.5x3.5cm in the outer upper quadrant extending into the central region and is 1cm away from the nearest deep resection margin. It is hard with greyish white cut surface and is surrounded by dense fibrosis. The rest of the breast appears predominanatly fibrofatty. Sectioning the axillary pad of fat reveals 5 lymph nodes, largest measuring 1.3x1x0.8cm with a greyish brown cut surface.
A) Nipple and areola 5 all (A1-A5)
B) Tumour 5 bits (B1-B5; B1 tumour with nearest deep resection margin)
C) Upper outer quadrant 1 bit (1 block)
D) Lower outer quadrant 1 bit (1 block)
E) Lower inner quadrant 1 bit (1 block)
F) Upper inner quadrant 1 bit (1 block)(B-F; The deep resection margin inked with india ink.)
G) 5 axillary lymph nodes 10 all (G1-G6; G1 and G2 from the same node)
NS/lv
MICRO :
A) Shows skin of nipple and areola, free of tumour.
B) Shows breast parenchyma infiltrated by a tumour arranged in tubules, nests, clusters and trabeculae displaying moderate pleomorphism, coarse nuclear chromatin, inconspicuous nucleoli and amphophilic cytoplasm. High grade DCIS is present. Lymphovascular invasion is evident. Perineurial invasion is not evident. Tumour is 1 cm from deep resection margin.
C-F) Shows breast parenchyma with no specific lesion.
G) Shows 5 lymph nodes, free of tumour.
IMPRESSION : Invasive ductal carcinoma, grade-II, right modified radical mastectomy.
Maximum size of the tumour is 4cm.
High grade DCIS present.
Lymphovascular invasion present.
Perineurial invasion, not evident.
Tumour is 1cm from the deep resection margin.
Skin of nipple and areola, free of tumour.
5 lymph nodes, free of tumour.
pT2N0Mx.

Fig. 3.Sample Pathology Report in English

D. Machine Translation of Pathology reports

MT systems are classified into General, Rule-based Example-based, Knowledge-based, Statistical and Hybrid systems. In the developed CDSS, machine translation of Breast cancer pathology reports is performed by Phrase-based translation approach. Machine translation is done from English to Tamil. English is syntactically a Subject-Verb-Object (SVO) language in which the word order is rigid and fixed while Tamil is a Subject-Object-Verb (SOV) language. Since it allows flexibility in word order it can be called "word-order free" language. In Tamil it is not mandatory that a sentence must have Subject, Object and Verb. A single word in Tamil can show the tense, action performed and the gender. In translating from English to Tamil, translation of gender-specific statements and ambiguous words are complex than in English. The MT process need not address this issue because all reports are of female patients and the Pathology report does not have any gender-specific mentions in its content. The complexity of translation due to ambiguous words also does not apply to translation by BEACON DSS since the textual content has standard medical terms and phrases which are locally and globally used by Pathologists. The machine translation process requires no human intervention, because it uses relevant domain resources.

Before translation, the system preprocesses the source text in English using NLP step. The Section Segregation divides the report into its constituent sections. The translation process is applied on individual sentences; hence the sentence splitting is performed next. The sentences are POS-tagged and stored. The PENN treebank tag set containing 36 tags is used for tagging. This is followed by splitting of each sentence for the translation process. Medical terms have phrases or group of words which cannot be split, in order to retain their medical meaning. Hence chunking is performed before translation. For example, a term "Lymphovascular invasion" is meaningful as a phrase than as individual word components. Case-sensitivity is handled at the preprocessing stage to avoid errors in Dictionary lookup during translation. Abbreviations are directly translated to their expanded Tamil equivalents while the non-translatable terms appear without changes in the translated output. The algorithm given below shows the steps in the translation process and Fig. 5 gives an example.

Algorithm

1. Retrieve and read a Pathology report in English
2. Preprocess the report contents
3. For Report 1 to n
4. For Sentence S 1 to m
   POS Tag and store in D
5. Split S into phrases
6. For each phrase P in S
   Apply Chunking
7. For each term n to 1, in the Chunk list
   If chunk-term Є Bilingual Dictionary
   Replace Chunk-term with term in Tamil
   Go to 6
   Endif
   End
8. Read pre corrected translation output in Tamil.
9. Apply Morphological Rules
10. Check terms that are out of context and replace with Synset
10. Print the translated Pathology report

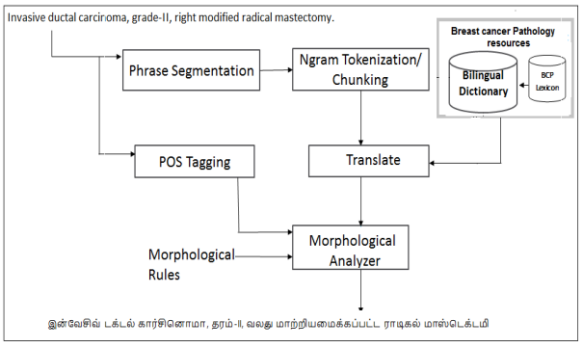


Fig. 4.Phrase-based Translation approach

The translated output is not initially syntactically correct, until Morphological analysis is applied to correct the inflated forms of a verb or noun. Hence, Morphological analysis is performed on the output to obtain perfect translation. The POS tagged report content was manually scrutinized several times to derive the rules. The Morphological rules refine the first-level translated output into an output form that adheres to grammatical rules in Tamil language. An example of such a corrected output is given below.

English: *Tumour is 1 cm from the deep resection margin.*
Translation before Morphological Analysis: கட்டி is 1 செ. மீ from the ஆழமான வெட்டு விளிம்பு.
Translation output after Morphological Analysis: கட்டி ஆழமான வெட்டு விளிம்பில் இருந்து 1 செ.மீ இல் உள்ளது.

E. Context-sensitivity through Synsets

A synonym set, or Synset is a group of synonyms. Synset for the domain were manually compiled using online resources and are used to correct terms that are out of context in the translated output.
For the term: Inflammatory, the Synonyms are as listed below:
Synonym1 : அழற்சி சார்ந்த
Synonym 2: வீக்கம் உண்டுபண்ணுகிற
Synonym 3: ஆத்திரமூட்டுகிற
Synonym 4: வீக்க அழுகல் .
Preferred term: வீக்கம் உண்டுபண்ணுகிற
The system displays the translated output and the Pathologist selects the terms that have been translated and are out of context. For the selected term, the system performs look-up on the existing synset repository and lists the terms. The medical expert chooses the appropriate term to be replaced in the output. Thus the initial translation is fine-tuned to have context-sensitive terms that represent the domain.

IV. RESULTS AND EVALUATION

A CDSS test run allowed the Risk Assessment Questionnaires to be filled by regional-language literate women. It provided clear understanding and involvement by the patients,

as the respondent could read and understand the contents. The CDSS risk assessment on General parameters is shown in Fig. 6.

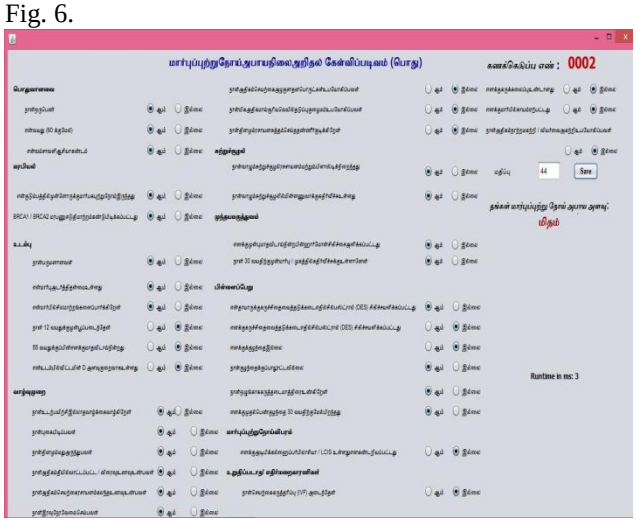


Fig. 5.*CDSS Risk Assessment output -General*

The most important patient-inclusive CDSS feature is the translation of Pathology report from English to Tamil. The Phrase-based translation of the report using custom-built translation resources for the domain, followed by fine-tuning of the translated output using Morphological analysis and Synset yielded good results. The Pathology report in Fig. 4, translated into Tamil is shown in Fig. 7

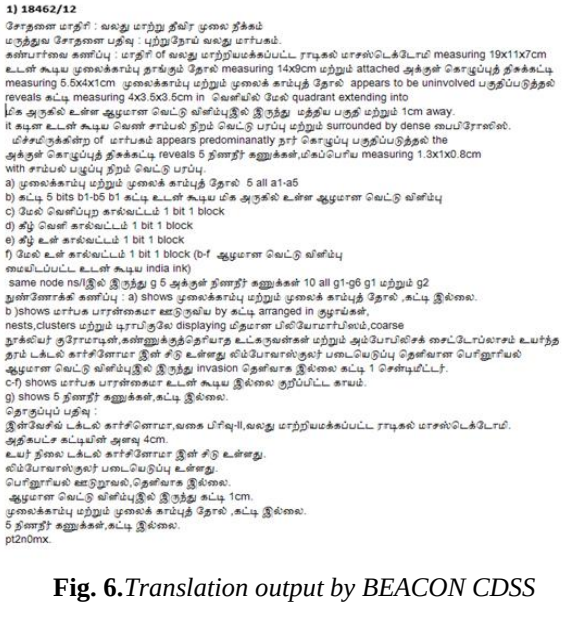


Fig. 6.*Translation output by BEACON CDSS*

The evaluation of the translation approach was done through comparison. The CDSS translation was compared with the translation of the reports using Google Translate output (Statistical MT and Neural MT), translation by English to Indian Language Machine Translation (EILMT) by Department of Information Technology, Government of India and manual scrutiny by the Language expert. Google Translate is a free online machine translation tool. In its earlier approach to translation using Statistical MT, it has limitations due to its inability to select the equivalent word in target language according to context, lack of sensitivity to syntax and medical terms. [29]. Google Translate has only 57.7% accuracy when used for medical phrase translations and should not be trusted for important medical communications. [30]. The BEACON CDSS provides correct and context-sensitive translation of the source text than Google Statistical MT, due to the use of appropriate translation resources, the translation approach and use of Synsets. The neural MT currently used by Google Translate yields better result than BEACON CDSS. The comparative translation output of the report section by Google Translate that used Statistical MT and in BEACON DSS are given in Fig. 8a and Fig.8b.

Fig. 8a.Google Translate (Statistical MT)

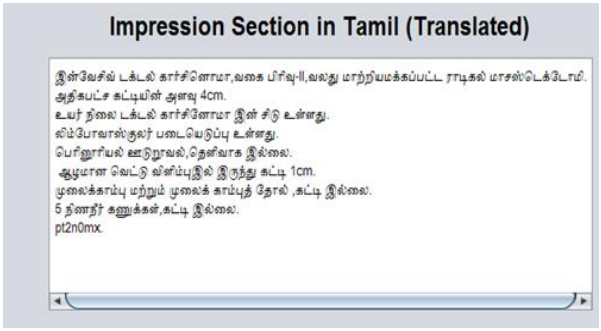


Fig. 7.b. Translation by BEACON CDSS

EILMT project translates from English to eight Indian languages; namely, Hindi, Bengali, Marathi, Urdu, Tamil, Oriya, Gujarati and Bodo in Tourism and Health domain. EILMT uses TAG based MT (Parser and Generator) and Statistical MT (SMT). The translation output of EILMT is given in Fig. 9. It is evident that the translation yields many sentences that have no proper meaning or structure.

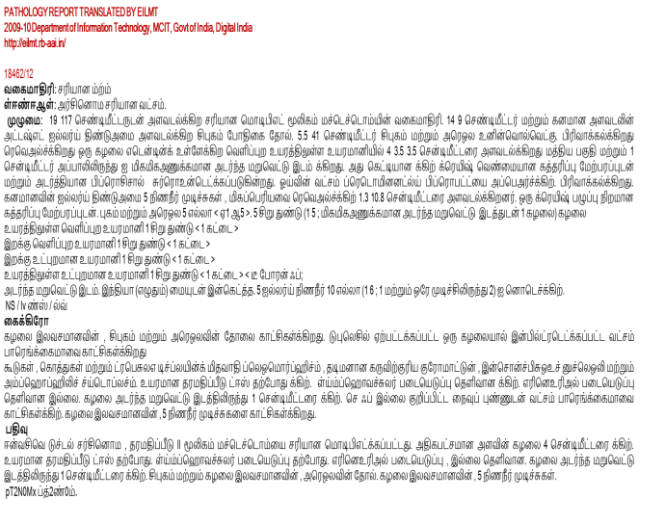
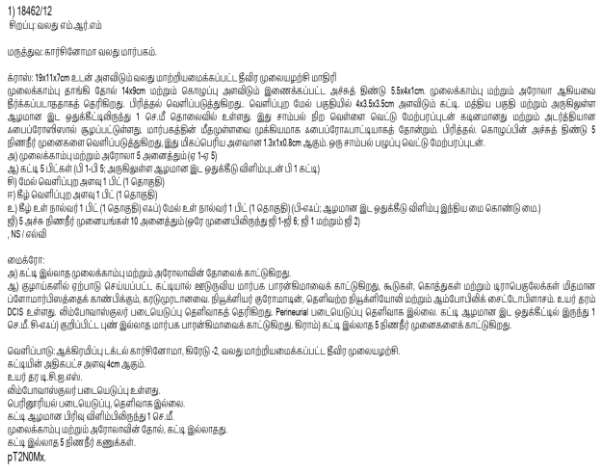


Fig. 8.Translation output by EILMT

BEACON CDSS translation is shown in Fig. 10. Comparing the three translation outputs in Fig. 8a. and 8b. 9 and 10, it is evident that BEACON CDSS performs better than EILMT and Google Translate (Statistical MT). The Neural MT BY Google Translate outperforms BEACON CDSS. Further evaluations were performed by a Tamil Language expert and a Medical expert. The criteria for grading by the experts is twofold. Firstly, the Tamil equivalent for the breast cancer pathology term must represent the medical aspect without loss of meaning and secondly the coinage of phrases must be syntactically and semantically correct from the target language perspective. The experts checked the correctness of translation and graded them as Good (G), Partially correct (PC) or Poor (P). Out of the 150 breast cancer pathology report translations, 142 translations were graded to be 'Good', 5 were graded to be 'Partially correct' and 3 were graded to be 'Poor'.


1) 18462/12
சிறப்பு: வலது எம்.ஆர்.எம்

மருத்துவ: கார்சினோமா வலது மார்பகம்.

கற்றல்: 19x11x7cm உடன் அளவிடும் வலது மாற்றியமைக்கப்பட்ட தீவிர முலையழற்சி மாஸ்டி
முலைக்காம்பு தாங்கி தோல் 14x9cm மற்றும் கொழுப்பு அளவிடும் இணைக்கப்பட்ட அச்சுத் திண்டு 5.5x4x1cm, முலைக்காம்பு மற்றும் அரோலா ஆகியவை நீக்கப்படாதகாத் தெரிகிறது. பிரித்தல் வெளிப்படுத்துகிறது. வெளிப்புற மேல் பகுதியில் 4x3.5x3.5cm அளவிடும் கட்டி. மத்திய பகுதி மற்றும் அருகிலுள்ள ஆழமான இட. ஒதுக்கீட்டிலிருந்து 1 செ.மீ தொலைவில் உள்ளது. இது சாம்பல் நிற வெள்ளை வெட்டு மேற்பரப்புடன் கடினமானது மற்றும் அடர்த்தியான கனப்பரேஷனஸ் குழப்பட்டுள்ளது. மார்பகத்தின் மீதமுள்ளவை முக்கியமாக கனப்பரேஷனட்டியாகத் தோன்றும். பிரித்தல். கொழுப்பின் அச்சுத் திண்டு 5 நிணநீர் முனைகளை வெளிப்படுத்துகிறது. இது மிகப்பெரிய அளவான 1.3x1x0.8cm ஆகும். ஒரு சாம்பல் பழுப்பு வெட்டு மேற்பரப்புடன்.
அ) முலைக்காம்பு மற்றும் அரோலா 5 அனைத்தும் (ஒ 1-ஒ 5)
ஆ) கட்டி 5 பிட்.கள் (பி 1-பி 5, அறிகுறியை ஆழமான இட. ஒதுக்கீடு விளிம்புடன் பி 1 கட்டி)
இ) மேல் வெளிப்புற அளவு 1 பிட். (1 தொகுதி)
ஈ) கீழ் வெளிப்புற அளவு 1 பிட். (1 தொகுதி)
உ) கீழ் உள் நால்வர் 1 பிட். (1 தொகுதி) (வெப்) மேல் உள் நால்வர் 1 பிட். (1 தொகுதி) (பி-வெப்: ஆழமான இட ஒதுக்கீடு விளிம்பு இந்திய மை கொண்டு மை.)
ஜி) 5 அச்ச நிணநீர் முனையங்கள் 10 அனைத்தும் (ஒரே முனையிலிருந்து ஜி 1-ஜி 6; ஜி 1 மற்றும் ஜி 2)
, NS / எல்லி

மைக்ரோ:
அ) கட்டி இல்லாத முலைக்காம்பு மற்றும் அரோலாவின் தோலைக் காட்டுகிறது.
ஆ) குழாய்களில் ஏற்பாடு செய்யப்பட்ட கட்டியால் ஊடுருவிய மார்பக பாரன்கிமாவைக் காட்டுகிறது. கூடுகள், கொத்துகள் மற்றும் டிராபெகுலேக்கள் மிதமான பலமாறுபிலத்தைக் காண்பிக்கும், கரடுமுரடானவை. நியூக்ளியர் குரோமாடின், தெளிவற்ற நியூக்ளியோலி மற்றும் ஆம்போபிலிக் சைட்டோபிளாசம் உயர் தரம் DCIS உள்ளது. விம்பேலாஸ்குலர் படையெடுப்பு தெளிவாகத் தெரிகிறது. Perineural படையெடுப்பு தெளிவாக இல்லை. கட்டி ஆழமான இட. ஒதுக்கீட்டில் இருந்து 1 செ.மீ. (வெப்) குறிப்பிட்ட புண் இல்லாத மார்பக பாரன்கிமாவைக் காட்டுகிறது. (ரோம்) கட்டி இல்லாத 5 நிணநீர் முனைகளைக் காட்டுகிறது.

வெளிப்பாடு: ஆக்கிரமிப்பு டக்டல் கார்சினோமா, கிரேடு -2, வலது மாற்றியமைக்கப்பட்ட தீவிர முலையழற்சி.
கட்டியின் அதிகபட்ச அளவு 4cm ஆகும்.
உயர் தர டி.சி.ஐ.எஸ்.
விம்பேலாஸ்குலர் படையெடுப்பு உள்ளது.
பெரினூரியல் படையெடுப்பு தெளிவாக இல்லை.
கட்டி ஆழமான பிரிவு விளிம்பிலிருந்து 1 செ.மீ.
முலைக்காம்பு மற்றும் அரோலாவின் தோல், கட்டி இல்லாதது.
கட்டி இல்லாத 5 நிணநீர் கணுக்கள்.
pT2N0Mx.

Fig. 9. Translation output by BEACON CDSS

V. CONCLUSION

Diagnosis of breast cancer is a life-altering and shocking discovery for a patient. A patient who is highly ignorant of the extent of the disease and treatment options will not be able to cope-up with the long years of treatment and care that follows. The need to bring together the Pathologist and the patient through better understanding and communication between the two is vital in breast cancer treatment and follow-up. Decision regarding surgery for the breast cancer patients is also an emotional and stressful decision. Sharing adequate information and collective decision by the Doctor and the patient enhances the understanding of the patient and eases the pain through a more co-operative decision-making process. The benefits of patient participation include: increased patient satisfaction and trust, reduced patients' anxiety and emotions, better understanding of personal requirements, improved health outcomes, enhanced quality of life, and delivery of more appropriate and cost effective services. Today, patient participation is regarded as a legal right of the patient as well as an International gold standard for healthcare systems. [14] Health professionals generally have a positive attitude toward patient participation, and consider this concept as a special privilege for themselves and the patients. Hence patient-oriented healthcare, based on opinions, needs, and preferences of patients must be adapted in India across patients of all strata of society. Incorporating "regional language" dimension as one of the features of CDSS and building translation resources namely the Lexicon and Bilingual dictionary (English to Tamil) for the breast cancer pathology domain, design of templates for Risk assessment based on General and Family History perspectives and translation of Pathology reports from English to Tamil, are the specific contributions of this work. The development of a CDSS with user services with regional language support is an initiative that the authors believe is first of its kind. The future scope of this work is to develop the actual system based on the prototype and extend the regional language features to other regional languages in India. Providing a patient-inclusive decision making system would not only encourage the CDSS use across many hospitals in the future but also improve health care outcomes in the considered breast cancer pathology domain.

ACKNOWLEDGMENT

The authors thank the Department of Pathology, Christian Medical College and Hospital, Vellore for providing them with the sample data for their study. Our special thanks to Dr. Marie Therese Manipadam and Dr. Gunadala Ishitha, Department of Pathology, CMC, Vellore for sharing their domain expertise. We also thank the Tamil language expert Dr. P. David Prabhakar, Head of the Department of Tamil, Madras Christian College, for the evaluation of Tamil translations. The contributions of Mr. Manikandan Kabali, and Mr. Karthik, students of MCA in the Department of Computer Science, Madras Christian College in developing the application, is also deeply appreciated.

REFERENCES

1. Sara Heath, 3 Best Practices for Shared Decision-Making in Healthcare: Shared decision-making in healthcare and improve patient health outcomes and support patient-centeredness during care encounters, https://patientengagementhit.com/news/3-best-practices-for-shared-decision-making-in-healthcare.
2. Prashant Mishra, Managing Director BMJ India & South Asia, Role of Evidence-Based Medicine in improving quality of healthcare, ETHealthWorld, October 30, 2018, https://health.economictimes.indiatimes.com/news/industry/role-of-evidence-based-medicine-in-improving-quality-of-healthcare/66428983
3. Ambigapathy R, Chia YC, Ng CJ. Patient involvement in decision making: a cross-sectional study in a Malaysian primary care clinic. BMJ Open 2016;6:e010063, doi:10.1136/bmjopen-2015010063.
4. Sridevi Seetharam and Renzo Zanotti, Patients' Perceptions on Healthcare Decision Making in Rural India: A Qualitative Study and Ethical Analysis, The Journal of Clinical Ethics, Volume 20, Number 2, pp. 1-9, 2009.
5. A M Stiggelbout, Shared decision making: really putting patients at the centre of healthcare, BMJ, 4 February 2012, Volume 344, pp. 28 - 31.
6. Involving people in their own health and care: Statutory guidance for clinical commissioning groups and NHS England, https://www.england.nhs.uk/wp-content/uploads/2017/04/ppp-involving-people-health-care-guidance.pdf
7. Ronak Sumbaly, N. Vishnusri. S. Jeyalatha, Diagnosis of Breast Cancer using Decision Tree Data Mining Technique, International Journal of Computer Applications (0975 – 8887), Volume 98– No.10, July 2014.
8. Spencer Robinson , Veronique Poirier, Sam Watson, Using Cancer Decision Support Tools to support the early diagnosis of cancer, Accelerate, Coordinate, Evaluate (ACE) Programme, An early diagnosis of cancer initiative supported by NHS England, Cancer Research UK and Macmillan Cancer Support.
9. M. Alamelumangai, B. Sathiyabhama, Personalized Care: A Clinical Decision Support System for Breast Cancer Screening Using Clustering and Classification, Elsevier's SSRN eLibrary – Journal of Information Systems & eBusiness Network – ISSN: 1556-5068.

10. Rajeev Chaudhry et.al, NLP-enabled decision support for cervical cancer screening and surveillance, September 30, 2014 to November 29, 2016
11. Ahmed M. Alaa, Kyeong H. Moon, William Hsu, Mihaela van der Schaar, Fellow, IEEE, ConfidentCare: A Clinical Decision Support System for Personalized Breast cancer Screening, https://arxiv.org/abs/1602.00374v1 [cs.LG], 2016.
12. Rajesh Kumar Sinha, Manohara Pai, M M Manohara Pai, Vidyasagar Mamidipudi, Vadhiraja Bm, A Novel Knowledge Base Decision Support System Model for Breast Cancer Treatment, April 2010, DOI: 10.4038/sljbmi.v1i2.1609
13. Henry NL, Bedard PL, DeMichele A, Standard and Genomic Tools for Decision Support in Breast Cancer Treatment, American Society of Clinical Oncology Educational Book. 2017;37:106-115, PMID:28561710, DOI:10.14694/EDBK_175617
14. Shaghayegh Vahdat, Leila Hamzehgardeshi, Somayeh Hessam, Zeinab Hamzehgardeshi, Patient Involvement in Health Care Decision Making: A Review, Iran Red Cres Med J. 2014 January; 16(1): e12454. DOI: 10.5812/ircmj.12454 Published online 2014 January 5.
15. Antony P J, 2013, "Machine Translation Approaches and Survey for Indian Languages", International Journal of Computational Linguistics and Chinese Language processing, Volume 18, No. 1, pp. 47-78.
16. Thenmozhi, D & Aravindan, Chandrabose. (2018). Ontology-based Tamil–English cross-lingual information retrieval system. Sādhanā. 43. 10.1007/s12046-018-0942-7.
17. Poornima C, Dhanalakhmi V, Anand Kumar M, Soman K P, 2011, "Rule-based Sentence Simplification for English to Tamil Machine Translation System", International Journal of Computer Applications, (0975-8887), Volume 20-No.3.
18. Harshwardhan R, Mridula Sara Augustine, K.P. Soman, 2011, "Phrase based English to Tamil Translation System by Concept Labeling using Translation memory", International Journal of Computer Applications (0975-8887), Volume 20-No.3.
19. Dhanesh N, 2016, "A Conceptual Framework for Automated English to Tamil Machine Translation System", International Journal for Trends in Engineering & Technology, Volume 16 Issue 1.
20. Saraswathi S, P. Kanivadhana, M. Anusiya, S. Sathiya, 2011, "Bilingual Translation System", International Journal on Computer Science and Engineering", Vol.3 No.3 pp. 1168-1174.
21. Robert A. Greenes, David W. Bates, Kensaku Kawamoto, Blackford Middleton, Jerome Osheroff, Yuval Shahar, Clinical decision support models and frameworks: Seeking to address research issues underlying implementation successes and failures, Journal of Biomedical Informatics, Volume 78,2018, Pages 134-143, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2017.12.005.
22. Stephanie Medlock, Jeremy C Wyatt, Vimla L Patel, Edward H Shortliffe, Ameen Abu-Hanna, Modeling information flows in clinical decision support: key insights for enhancing system effectiveness, Journal of the American Medical Informatics Association, Volume 23, Issue 5, September 2016, Pages 1001–1006, https://doi.org/10.1093/jamia/ocv177.
23. http://dictionary.tamilcube.com/
24. https://glosbe.com
25. http://siddhadreams.blogspot.com/2008/10/tamil-medical-dictionary.html
26. ttp://www.tamilvu.org/library/technical_glossary/html/techindex.htm
27. www.translate.google.com
28. www.shabdkosh.com
29. Sheppard F, 2011, "Medical writing in English: The problem with Google Translate" La Presse Médicale, Volume 40, no. 6, pages 565-566. Doi : 10.1016/j.lpm.2011.02.024
30. Sumant Patil, Patrick Davies, Use of Google Translate in medical communication: evaluation of accuracy, BMJ 2014; 349 doi: https://doi.org/10.1136/bmj.g7392 (Published 15 December 2014)

AUTHORS PROFILE

Mrs. G. Johanna Johnsi Rani is currently working as Assistant Professor in the Department of Computer Science, Madras Christian College (Autonomous) in Chennai with 32 years of Teaching experience and 5 years of Research experience. The area of research is Natural Language Processing. She has 8 publications in International Journals and has presented research papers in 12 International Conferences.

Dr. (Mrs.) Gladis Christopher is the Principal of Bharathi Women's College in Chennai since 2018. Earlier she headed the PG and Research Department of Computer Science, Presidency College, Chennai. She has guided more than fifteen M.Phil and Ph.D Research scholars and has been an active researcher in the areas of Neural Networks, Image Processing, and Data Mining.

Dr. Joy John Mammen is currently the Professor and Head of the Department of Transfusion Medicine & Immunohaematology in CMC. His main research work focuses on Laboratory and Healthcare Informatics, Digital Imaging in Pathology, Error reduction & Tracking, Lab Decision Support, Patient care and National External Quality Assessment.