[1]Kuwali Talukdar

[2]Shikhar Kumar Sarma

Enabling Natural Language Processing and AI Research in Low-Resource Languages: Development and Description of an Assamese UPoS Tagged Dataset

Journal of Electrical Systems

Abstract: - This paper describes in detail the Universal Parts of Speech (UPoS) tagged dataset for the Assamese language. PoS tagged dataset in a language is crucial for experimenting and creating resources for various Natural Language Processing (NLP) and AI research. With the growing usage of Universal Dependency standards, tagged dataset with Universal PoS tags are becoming very much essential for contemporary experiments in the NLP community. NLP research in Assamese, and Indo-Aryan language, is relatively new, and the language is considered a Low Resource language. The dataset of UPoS tagged Assamese text is created with an aim of contributing towards enriching this low resource language for NLP and AI tasks. The size of the dataset is 283506 tokens of Assamese vocabulary, against total 20280 sentences, tagged with 17 standard UPoS tags of core lexical categories. The raw data are taken from an open-source corpus originally tagged with BIS tagset. The original size of 453457 tokens against 29504 sentences, after subjected to data filtering, was reduced to this clean resource of 283506 tokens. Lexical categories mapping is done with linguistic expertise, from BIS to UPoS tagsets. Mapped pattern was used for a first-level conversion of BIS tags to UPoS tags. Linguistic validation is also performed with linguistic experts and inter annotator agreement/disagreements were recorded. Second level validation resulted in deciding on the agreement, producing the final version of the dataset. This Assamese UPoS tagged dataset is the first of its kind with UPoS annotations and shall serve a wider Assamese NLP research community for model training using Machine Learning/Deep Learning Techniques.

Keywords: Assamese; PoS Tags; UPoS Tags; NLP; Assamese UPoS Tagset

Dataset: https://gauhati.ac.in/academic/technology/information-technology

Dataset License: license under which the dataset is made available (CC0, CC-BY, CC-BY-SA, CC-BY-NC, etc.)

I. INTRODUCTION

Parts of Speech tagging is an essential stage in Natural Language Processing (NLP), and it compliments and supports performances of many NLP applications, including Machine Translation, Language Generation, Question-Answering, Sentiment Analysis etc. Tagging uses a standard set of tags so that transparency and adaptability of heterogeneous processes and systems could be achieved. Natural Language understanding and generation research are basically dependent on resources, and standardization of such resources have been a matter of concern in the NLP community. Text resources with specific characteristics tagging have been used for a range of processes across languages. Such tagging ranges from PoS tagging to Named Entity, Multiword tagging etc., PoS being the dominant and most common one. As new languages are being added in the NLP research, lesser used languages start facing various challenges during the initial period, particularly regarding the resources for baseline research. Thereby creating demands for resources essentially required to start baseline research. Resource poor or low resource languages hence require specific attention in creating such resources. This includes corpus, parallel corpus, and tagged corpus. PoS tagging, as an integral annotation for language processing task, is therefore considered as an initial phase research task, and facilitated with Machine Learning and Deep Learning techniques, creates ample scope and room for a range of experiments. Deep Learning is always resource hungry, and without having a sizable

[1]Department of Information Technology, Gauhati University, India 781014; kuwalitalukdar@gmail.com

[2]Department of Information Technology, Gauhati University, India 781014; sks001@gmail.com

*Correspondence: kuwalitalukdar@gmail.com (K.T.); sks001@gmail.com (S.K.S.); Tel.: +91-8011003093

dataset, it cannot outperform lesser performing systems. The challenge of creating a sizable, as well as good quality dataset, is therefore essential. This could be achieved by starting with a base of standardized dataset creating manually with process intervention, and using this for experimenting with, modelling ML systems to facilitate machine tagging, and then subjected to the validation for having a purified dataset.

PoS tagged corpus, as a basic requirement for NLP tasks is always a sought for resource for the research community. Researchers in a plan to undertake low resource language processing tasks get the first hit of non-availability of good quality PoS tagged standardized text corpus. This is true in academic level research capacity building, as well as application development scenarios. A good quality PoS tagged dataset should essentially be an inclusive corpus with presence of all types of lexical categories having sufficient frequency of tokens and is expected to include a variety of linguistic phenomenon including presence of tokens with multiple senses used with multiple lexical categories, sufficiently representing the complexities of word morphologies and language syntax. Also, such a PoS tagged dataset should be with standardized tagged set, incorporating simple and homogeneous annotation conventions.

II. REVIEW

The evolution PoS tagset is seen historically for English language, with Brown Corpus using a standardized PoS tagset long back in 1979. In the Brown corpus, 82 different tags were used, and they termed these tags as word-classes, containing parts of speech, punctuation marks, and few inflectional morphemes [1], with a declared purpose of the tagged corpus for facilitating automatic or semi-automatic syntactic analysis. The Penn Treebank, a large corpus of over 4.5 million American English words was completely annotated with the designed Penn Treebank PoS tagset, which was far brief than the Brown Corpus PoS tagset. The downsized Penn Treebank tagset contains only 36 PoS tags and 12 other tags for punctuations and currency symbols [2], modifying the Brown Corpus tagset eliminating the lexical and syntactic redundancies. Since the pioneering Brown Corpus extensive tagset and the modified Penn Treebank tagset, other customized tagsets have been started evolving, extending also to language specific tagsets other than English. IBM Watson project used a tagset of 42 PoS tags for English, and a simplified tagset of only 16 tags specifically customized for non-English documents [3].

An ambitious project of having very large PoS tagged corpus was the British National Corpus (BNC) of 100 million words tagged with BNC PoS tagset. Here they used two different tagsets, one, the BNC Basic Tagset, and the other, the BNC enriched tagset. The BNC basic tagset contained 61 grammatical categories, whereas the BNC enriched tagset contained a larger set of word labels of 139 different defined categories, providing more detailed grammatical distinctions [4][5].

India is a multilingual country, having languages of various families used in different geographies. These languages have different common, as well as individual linguistic characteristics. Although evolving a common tagset for all Indian Languages has been a matter of concern and research, right from the initial days of Indian Languages NLP, with continuous and considerable coordinated efforts, a simple, common, supplemented with inclusive individual linguistics features taping, is yet to be framed. Discrete tagsets have evolved over years, starting from ILMT (Indian Language Machine Translation) project tagset, to the IIIT Hyderabad, and CIIL Mysore developed documents, and they have been seen covering a wide range of Indian languages.

A unified PoS tagset for Indian languages has been designed by TDIL (Technology Development programme under the Ministry of Communication and Information Technology, Govt. of India. [6]-[8]. Here 45 different lexical annotations are defined under 11 top level PoS categories. This unified tagset also contains language specific customized tagset for various Indian languages- Hindi, Punjabi, Urdu, Gujarati, Bengali, Telegu, Kannada, Malayalam, Tamil, Marathi, Konkani, and Maithili.

A common framework for Indian languages PoS tagsets has also been worked out separately by a group of NLP researchers led at Microsoft India Research Lab [9]-[12]. Known as IL-POSTS, this hierarchical framework has been designed to accommodate language specific requirements and claimed to interoperable and flexible for among the Indian languages. This tagset contains 11 categories that are mandatory for any language, followed by different Types against each category, and 18 attributes. While they categorised Types as recommended for all the languages, Attributes are defined as optional, thereby adding flexibility to have a subset of the main tagset.

More recently the Bureau of Indian Standard (BIS) published Indian language PoS tagset is a comprehensive tagset, evolved into a common set for all Indian languages, supplemented with language specific tagset for few languages [13]-[15]. This standard facilitates a superset applicable for all Indian languages, and defines 11categories, with annotations for 45 subcategories. The same tagset is also customized with language specific tagsets.

BIS tagset is observed to be the de-facto PoS tagset for most of the PoS tagging works for Indian languages in recent days, and NLP experiments across the Indian languages nowadays use resources tagged with BIS tagsets.

A more recent PoS tagset standard evolved with the Universal Dependency treebank project is the Universal Parts of Speech (UPoS) tagset. This tagset consists of 17 universal PoS tags defined as core categories [16]. Here, for other lexical and grammatical properties of words, universal features are used. Features could be evolved and defined language specific, thereby giving flexibility to adopt required annotations to accommodate languages with specific lexical information.

Assamese NLP works for a variety of experiments are found to use various resources, few standardized, but mostly discrete nonstandard, used specific to the task. The language is a member of Indo-Aryan family of languages, and is the official state language of Assam, the northeastern state of India. It is one of the scheduled languages of India and spoken by approximately 15 million native speakers. Digital penetration in India has been very swift, and as most of population use their mother tongue or first language for communication, work, business etc., demand for local language digital applications is very huge, resulting a push for local language NLP research and development works. Assamese, relatively new in NLP research has also got a swift expansion in terms of research and development base regarding different NLP. And this has created a new requirement for standardized dataset. Although many research activities could be traced in various NLP aspects, like annotations, machine translation, words sense disambiguation, corpus building etc. [17], but publication/release of standardized dataset with proper validation of quality is rarely seen.

III. SUMMARY

The current work concentrates on creating a standard dataset for Assamese language, with a sizable amount of text annotated with UPoS tags. As PoS tagging is an important stage for almost all the NLP research pipelines, the requirement of PoS tagged dataset is critical to the performance of system. The emergence of Machine Learning and Deep Learning as the most effective techniques for NLP tasks has created the inevitable demand for quality and sizable dataset. Here in this paper, we detail the dataset created for Assamese language, as an UPoS annotated text dataset. The Assamese UPoS tagged dataset is created through an evolutionary model stating from considering BIS tagged text corpus of 29504 sentences, with 453457 tokens. To create the UPoS tagged dataset, a conversion table already created with mapping of BIS tags to the 17 core UPoS tags is used. A customized python script is run over the BIS tagged sentences for the first round of replacement with UPoS tags. Filtering of the raw file is done to clean the data, so that the resultant dataset is at the expected quality level. The final data set is a reduced set after removing unwanted tokens and sentences, with a data size of 20280 sentences and 283506 tokens.

This dataset of Assamese text annotated with UPoS tags is of gold standard, as phases of linguistic validation have been performed. Filtering for removing unmatched segments, blank lines, and non-uniform non-standard text have contributed to having a cleaned dataset of highest quality. This dataset is the first ever resource on Assamese language UPoS tagged resources and shall serve a wider research community undertaking experimentation on PoS tagger design for Assamese language using ML/DL techniques. Models could be trained using this standard dataset with contemporary ML/DL techniques of RNN/GRU/BiLSTM etc., integrated with techniques of word embedding, for a well performed UPoS tagger.

IV. DATA DESCRIPTION

This dataset comprises of Assamese text with each token annotated with the corresponding lexical categories. The annotations use the UPoS standard annotations against UPoS defined 17 core categories. These 17 UPoS core categories are shown in Table 1.

Table 1: UPoS Core Categories

Open class words		**Closed class words**		**Other**	
Tags	**PoS Categories**	**Tags**	**PoS Categories**	**Tags**	**PoS Categories**
ADJ	adjective	ADP	adposition	PUNCT	punctuation
ADV	adverb	AUX	auxiliary	SYM	symbol
INTJ	interjection	CCONJ	coordinating conjunction	X	other
NOUN	noun	DET	determiner		
PROPN	proper noun	NUM	numeral		
VERB	verb	PART	particle		
		PRON	pronoun		
		SCONJ	subordinating conjunction		

The resource is in a text file format with the filename and extension as GUIT-AsUPoS-TaggedDataset.txt. The file name signifies as follows: GUIT is the Organization Identifier for Gauhati University Information Technology Department, where the dataset is created. As is the IPA nomenclature for Assamese language, followed by the nature of Tagset used-Universal Parts of Speech Tagset (UPoS). The dataset file contains tagged sentences as the primary distinguishable entity. These are arranged as lines, indicated by line numbers.

Table 2: Example part of the dataset.

গাঁত\N_NOUN{NN}খান্দিবলৈ\N_NOUN{NN}পাৰাৰ\RD_X{RDF}টিলাৰ\RD_X{RDF}
বা\CC_CCONJট্রেক্টৰ\N_NOUN{NN}চালিত\ADJগাত\N_NOUN{NN}খন্দা\V_VERB{VM}
যন্ত্ৰ\N_NOUN{NN}ব্যৱহাৰ\N_NOUN{NN}কৰা\V_VERB{VM} হয়\V_AUX ।\RD_PUNCT
ফলমুলৰ\N_NOUN{NN}খেতিত\N_NOUN{NN}ট্রেক্টৰ\N_NOUN{NN}বা\CC_SCONJ
পাৰাৰ\N_NOUN{NN}টিলাৰ\N_NOUN{NN}চালিত\ADJচহ\N_NOUN{NN}কৰা\V_AUX
সঁজুলি\N_NOUN{NN}ব্যৱহাৰ\N_NOUN{NN}কৰা\V_AUX হয়\V_AUX ।\RD_PUNCT
ফলমুলৰ\N_NOUN{NN}খেতিত\N_NOUN{NN}বনবাত\N_NOUN{NN}আদি\ADV
নিৰাবলৈ\N_NOUN{NN}যন্ত্ৰ\N_NOUN{NN}চালিত\N_NOUN{NN}নাঙল\N_NOUN{NN}ব্যৱহাৰ\N_NOUN{NN}কৰা\V_VERB{VM} হয়\V_AUX ।\RD_PUNCT শাক\N_NOUN{NN}
পাচলিৰ\N_NOUN{NN}গুটি\N_NOUN{NN}দিয়া\V_AUXযন্ত্ৰও\N_NOUN{NN}
উদ্ভাৱন\N_NOUN{NN}হৈছে\ADJ ।\RD_PUNCT গছৰ\N_NOUN{NN}ডাল\N_NOUN{NN}
আদি\ADVচিকুনাবলৈ\N_NOUN{NN}ব্যৱহাৰ\N_NOUN{NN}কৰা\V_AUXযন্ত্ৰও\N_NOUN{NN}
অসমত\N_PROPN ব্যৱহাৰ\N_NOUN{NN}কৰা\V_VERB{VM} হয়\V_AUX।\RD_PUNCT
উদ্যান\N_NOUN{NN}শস্যত\N_NOUN{NN}জলসিঞ্চনৰ\N_NOUN{NN}বাবে\X{PSP} সুক্ষ্ম\ADJ
জলসিঞ্চনৰ\N_NOUN{NN}ব্যৱস্থাও\N_NOUN{NN}কৰা\V_AUX হয়\V_AUX
।\RD_PUNCTজলসিঞ্চনৰ\N_NOUN{NN}ক্ষেত্ৰত\N_NOUN{NN}ড্রিপক\N_PROPN
স্প্রিকলাৰ\N_NOUN{Nloc}জলসিঞ্চন\N_NOUN{NN}ব্যৱস্থা\N_NOUN{NN} অতি\RP_PART{INTF}
উত্তম\ADJ ।\RD_PUNCT

The Assamese tokens are followed by the PoS annotations in the syntax-

"ASSAMESE_WORD\ANNOTATION".

The PoS annotation convention used is as follows-

-\XXX_YYY{ZZZ}

Where—

XXX denotes label from the BIS tagset;

YYY is the UPoS tag from the UPoS Core Categories, and

ZZZ is the subcategory signifying the lexical features.

The descriptive metadata of the dataset are as follows:

- Name of the dataset: Assamese UPoS tagged text dataset.
- Dataset File name: GUIT-AsUPoS-TaggedDataset.txt
- File format: text file with .txt file extension.
- Total number of Sentences: 20280
- Total Number of tagged tokens (words): 283506
- Number of tags: 17 Core categories
- PoS tagset: UPoS tagset as defined in the Universal Dependencies (UD)

Statistical patterns of the tokens as per UPoS annotation categories are also extracted. This is presented in Table 3.

Table 3: Frequency of tokens in the dataset as per lexical categories

Tags	PoS Categories	Frequency in the dataset	Frequency in the dataset as a %age of total tokens
Open class words			
ADJ	adjective	21307	7.52%
ADV	adverb	425	0.15%
INTJ	interjection	35	0.01%
NOUN	noun	117817	41.56%
PROPN	proper noun	16418	5.79%
VERB	verb	24944	8.80%
Closed class words			
ADP	adposition	4531	1.60%
AUX	auxiliary	12944	4.57%
CCONJ	coordinating conjunction	10423	3.68%
DET	determiner	6589	2.32%
NUM	numeral	11906	4.20%

PART	particle	3540	1.25%
PRON	pronoun	6082	2.15%
SCONJ	subordinating conjunction	2394	0.84%
Other			
PUNCT	punctuation	38315	13.51%
SYM	symbol	1222	0.43%
X	other	4615	1.63%

V. METHODS

The dataset is created with automatic tagging of UPoS annotations against the BIS tagged Assamese text, and then following a validation process. For mapping of BIS annotation conventions to the UPoS tags, the conversion mapping as shown in Table 4 is used. The initial BIS tags annotated Assamese text of approximately 30000 sentences volume has been subjected to data cleaning process. The data cleaning phase consist of filtering for the following parameters:

1. Removal of unwanted sentences: Sentences of abnormal lengths are detected and removed, assuming that these chunks may not be exact sentences, rather may be combined sentences, and shall not represent language syntax. As PoS is a sequential pattern following the language syntax of formation of well-formed sentence, such patterns may not contribute to the allowable sequence pattern.
2. Removal of unmatched sequences: Mis-matched sequences of tokens vs tags are detected and removed before creating the dataset. This is because this will result in chaotic mapping of BIS tag sequences to UPoS tag sequences and contributing to noisy dataset.
3. Foreign token removal: Tokens which are not in Assamese script are considered as foreign tokens and hence were removed. As the dataset is pure Assamese UPoS tagged dataset, the entities allowable in Assamese text are only considered to be present in the dataset, and hence this filtering is performed.

Table 4. UPoS-BIS tags mapping

UPoS Core Categories	**BIS Categories**
1. Noun (NOUN)	**Noun (N)**
	Common Noun (NN)
	Verbal Noun (NNV)
	Nloc (NST)
2. Proper noun (PROPN)	**Proper Noun (NNP)**
3. Verb (VERB)	**Verv (V)**
	Verb Main (VM)
	Verb finite (VF)
	Verb non-finite (VNF)
	Verb infinite (VINF)
	Gerund (VNG)
	Verbal (VN)
4. Pronoun (PRON)	**Pronoun (PR)**
	Personal Pronouns (PRP)

	Reflexive Pronoun (PRF)
	Relative Pronoun (PRL)
	Reciprocal Pronoun (PRC)
	Interrogative Pronoun/WH-word (PRQ)
	Indefinite Pronoun (PRI)
5. Adjective (ADJ)	**Adjective (JJ)**
6. Adverb (ADV)	**Adverb (RB)**
7. Adposition (ADP)	**Postposition (PSP)**
8. Auxiliary Verb (AUX)	**Auxiliary Verb (VAUX)**
	Finite (VF)
	Non finite (VNF)
	Infinitive (VINF)
	Participle Noun (VNP)
9. Coordinating conjunction (CCONJ)	**Conjunction (CC)**
	Coordinator (CCD)
10. Subordinating conjunction (SCONJ)	Subordinator (CCS)
11. Interjection (INTJ)	Interjection (INJ)
12. Determiner (DET)	**Demonstrative (DM)**
	Deictic (DMD)
	Relative (DMR)
	Wh-word (DMQ)
	Indefinite (DMI)
13. Numeral (NUM)	**Quantifiers (QT)**
	General (QTF)
	Cardinal (QTC)
	Ordinal (QTO)
14. Particle (PART)	**Particle (RP)**
	Particle Default (RPD)
	Classifier (CL)
	Intensifier (INTF)
	Negation (NEG)
15. Punctuation (PUNCT)	Punctuation (PUNC)
16. Symbol (SYM)	Symbol (SYM)
17. Others(X)	Residuals (RD)
	Foreign word (RDF)
	Unknown (UNK)
	Echowords (ECH)

The filtered text then consists of 20280 sentences and contains 283506 tokens. UPoS tags as defined in the UD UPoS tagset are then used for putting corresponding annotations against the tokens. The resultant UPoS tagged text corpus of 283506 tokens in size is thereafter subjected to linguistic validation. Manual cross validation is done on 40% of the total sentences with double blind reviews, and then inter-annotators agreements are observed. Errors found are corrected by individual validators, and then disagreements during blind validations are brainstormed for arriving at agreements. Error analysis and inter-annotators agreements/disagreements are shown in table 5. The error-corrected and modified resource is then considered as the final version of the Assamese UPoS tagged dataset.

VI. CONCLUSION

As UPoS is expanding as a universal standard of tagset for lexical category annotations across the languages, and designed to be language independent, UPoS tagged Assamese text dataset shall act as a valuable resource for Assamese NLP research community. Assamese is a rapidly growing language in digital sphere, with a sizable base of research groups contributing towards fundamental as well as application-oriented tasks. This dataset shall facilitate the Assamese NLP community, especially the research students for performing experimentations with diversed ML/DL techniques. Moreover, the system trained with this dataset could, in turn, evolve as automatic UPoS tagger yielding larger tagged resources. The dataset is made available on the website https://gauhati.ac.in/academic/technology/information-technology.

REFERENCES

[1] Francis, W. N. and Kucera, H. Brown Corpus Manual., Department of Linguistics, Brown University, Providence, Rhode Island, US (1979).

[2] Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large, annotated corpus of English: the penn treebank. Comput. Linguist. 19, 2 (June 1993), 313–330.

[3] https://www.ibm.com/docs/en/wca/3.5.0?topic=analytics-part-speech-tag-sets

[4] http://www.natcorp.ox.ac.uk/docs/gramtag.html

[5] TDIL (Technology Development Indian Languages) programme, Ministry of Communication and Information Technology, Govt. of India, Unified Parts of Speech (POS) Standard in Indian Languages, https://tdil-dc.in/tdildcMain/articles/134692Draft%20POS%20Tag%20standard.pdf

[6] Baskaran Sankaran, Kalika Bali, Monojit Choudhury, Tanmoy Bhattacharya, Pushpak Bhattacharyya, Girish Nath Jha, S. Rajendran, K. Saravanan, L. Sobha, and K.V. Subbarao. 2008. A Common Parts-of-Speech Tagset Framework for Indian Languages. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. European Language Resources Association (ELRA).

[7] Bureau of Indian Standards. (2021) “Linguistic Resources-POS Tag Set for Indian Languages-Guidelines for Designing Tagsets and Specification.” www.bis.gov.in, www.standardsbis.in

[8] Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. Computational Linguistics, 47(2):255–308.

[9] A.K. Barman, J. Sarmah and S. K. Sarma, "POS Tagging of Assamese Language and Performance Analysis of CRF++ and fnTBL Approaches," 2013 UKSim 15th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2013, pp. 476-479, doi: 10.1109/UKSim.2013.91.

[10] Anup Barman, Jumi Sarmah, and Shikhar Sarma. 2014. Assamese WordNet based Quality Enhancement of Bilingual Machine Translation System. In Proceedings of the Seventh Global Wordnet Conference, pages 256–261, Tartu, Estonia. University of Tartu Press.

[11] Baruah, K.K., Das, P., Hannan, A., & Sarma, S.K. (2014). Assamese-English Bilingual Machine Translation. ArXiv, abs/1407.2019.

[12] Sarmah, J., & Sarma, S.K. (2016). Survey on Word Sense Disambiguation: An Initiative towards an Indo-Aryan Language. International Journal of Engineering and Manufacturing (IJEM), Vol.6, No.3, pp.37-52, 2016.DOI: 10.5815/ijem.2016.03.04

[13] Shikhar Kr. Sarma, Himadri Bharali, Ambeswar Gogoi, Ratul Deka, and Anup Kr. Barman. 2012. A Structured Approach for Building Assamese Corpus: Insights, Applications and Challenges. In Proceedings of the 10th Workshop on Asian Language Resources, pages 21–28, Mumbai, India. The COLING 2012 Organizing Committee.

[14] Lakshmi Priya. G, Anjali J Nair (2024). Energy management strategy for fuel cell hybrid electric vehicle. IJRDO - Journal of Electrical and Electronics Engineering; 8(1): 1-11.

[15] Ayush kale (2021). Optimization of pulse forming networks. IJRDO - Journal of Electrical and Electronics Engineering; 7(7): 1-13.

[16] Truong Tuan Anh, Dao Duy Yen (2020). Some solutions for online monitoring the vibration of transformers and the application to identify the state of the transformers. IJRDO - Journal of Electrical and Electronics Engineering; 6(6): 1-12.

[17] Mandeep Kumar Sinha, Dr Ravindra M, (2020). Elderly Assistant Based on Facial Emotion and Posture Analysis. IJRDO - Journal of Electrical and Electronics Engineering; 6(6): 13-20.