Deep Learning Models of LSTM-Ann and Bilstm-ANN for Classification Accuracy

M. Srisankar
Research Scholar,
Department of computer science
Government Arts College, Udumalpet

K.P. Lochanambal, PhD
Assistant Professor
Department of computer science
Government Arts College, Udumalpet

ABSTRACT
In recent years, there has been an exponential growth in the number of complex documents and texts that require a deeper understanding of machine learning methods to be able to accurately classify texts in many applications. Many machine learning approaches have achieved surpassing results in natural language processing. The success of these learning algorithms relies on their capacity to understand complex models and non-linear relationships within data. However, finding suitable structures, architectures, and techniques for text classification is a challenge for researchers This paper has illustrated the deep integration of BiLSTM-ANN (Fully Connected Neural Network) and LSTM-ANN and manifested how these integration methods are performing better than single BiLSTM, LSTM and ANN models is discussed. This overview covers different classification methods, Sentiment analysis work, and existing algorithms and techniques, and evaluations methods. Finally, the limitations of each technique and their application in real-world problems are discussed.

Keywords
Tokenization, LSTM-ANN, Bi-LSTM-ANN

1. INTRODUCTION
With the progression of this new technology of collaborations, the innovation is getting progressively custom designed.Chatbots, computerized response structures, and recommender systems are increasingly becoming actually conversational, studying our propensities and characters or even growing characters of their personal. There are a number of profoundly beneficial applications of artificial Intelligence. This development is empowering training and revolutionizing individualized gaining knowledge of aid for increasing quantities of teachers and students. System mastering programs are converting the sector through advancing almost all the sectors, such as hospital treatment administrations, transportation, automation, different commercial production machine, and so forth. One of the critical troubles with algorithms like LSTMs is they are liable to overfitting. Also, it's far hard to apply the dropout calculation to clear up the overfitting difficulty. Moreover, LSTM is slower in evaluation to the call for the type mission. This is because of the successive operations within the LSTM layers. LSTM requires consecutive contributions to figure out the hidden layer weights iteratively, which makes it a chunk phlegmatic in terms of performance than the other neural networks.Even though BiLSTM plays higher than the LSTM, it has limitations too be able to tackle the ones limitations, content classification is the way toward grouping the contents and allotting labels to not unusual literature sources by using some predetermined set of categories algorithmically [1]. In this exploration, we have utilized principal recurrent neural networks which include LSTM and BiLSTM, and deeply incorporated them to make crossover fashions with synthetic neural networks for classifying the substance from new sources. Recurrent Neural networks are able to studying order dependence in succession forecast problems and are utilized extensively in language processing applications. each LSTM and BiLSTM are assimilated with every other deep ANN version with a view to make two higher models than the person development of LSTM and BiLSTM fashions. In addition, we examine and display that the hybrid model offers away most advantageous exactness and characterization end result over the single LSTM and BiLSTM fashions We have compared the BiLSTM -ANN and LSTM-ANN with the single deep development of BiLSTM andLSTM, respectively,

2. RELATED WORK
Hidayatullah A.F. et al. in [2] have proposed a model for adult content arrangement by utilizing Long Short-Term Memory (LSTM) neural network to characterize adult and non-adult content. In [3], Wang Y. et al. suggested a model that incorporates an attention-based Long Short-Term Memory network for aspect-level emotion or sentiment arrangement. The system can focus on various pieces of a sentence when various perspectives are taken as information. Nowak J. et al. showed how to order text utilizing the LSTM model and its modification. Further, the authors presented the predominance of this strategy over different calculations for text classification in [4]. Kumar A. et al. presented a multi-modular way to deal with recognizing debacle-related educational substance from the Twitter streams utilizing text and pictures together. Their methodology was based on LSTM and VGG-16 networks that show critical improvement in the exhibition, as obvious from the approval result on seven distinctive catastrophe-related datasets [5]. Yildirim O. et al. executed a convolutional auto-encoder (CAE) based nonlinear pressure structure to decrease the sign size of arrhythmic beats. LSTM classifiers are utilized to perceive arrhythmias utilizing ECG highlights naturally [6]. Chen T. et al. claimed to have applied a neural network-based arrangement model to characterize obstinate sentences into three classes as per the number of targets that showed up in a sentence. The authors then took care of Each type of sentence into a one-dimensional convolutional neural organization independently for supposition order [7]. Sun Q. et al. considered the spam job posting identification as the objective issue and assemble a nonexclusive machine learning pipeline for multi-lingual spam locations. The fundamental parts of their research are feature generation and information relocation through transfer learning [8]. Dumais S. et al. in [9] investigated the utilization of progressive design for grouping an enormous, heterogeneous assortment of web content. The authors have used a support vector machine (SVM) classifier, which has been demonstrated to be productive and compelling for characterization, yet not recently investigated with regards

to various hierarchical classifications. Lytvyn V. et al. in [10] explored the issue of mechanized advancement of fundamental ontology and further proposed an algorithm for extracting information from regular content. Karim F.et al. proposed the enlargement of convolutional networks with LSTM sub-modules for time sequence classification. The authors have fundamentally improved the presentation of convolutional networks with a nominal expansion in model measure and need insignificant preprocessing of the informational collection [11]. Kowsher, Md et al. showed the methodology of information extraction system from human names and expressed the better performance of LSTM based model [12]. Unlike the aforementioned research, we have fused sequential model BiLSTM and the ANN to get better accuracy in classification tasks.In contrast to this research, every one of the papers addressed above shares one common flaw, practically speaking. The authors have not looked for an original technique by incorporating at least two algorithms to change itinto a solitary hybrid model. Their strategies were more inclined to fall into traditional acts of data characterization or other machine learning applications. Though a portion of the authors had great accomplishment in elevating the exhibition through novel preprocessing strategies, they focused on only the application and execution of thealgorithm such as classification of adult contents, web information, and ECG data. Whereas, in this study, we are not just zeroing in on another AI application yet in addition anticipating upgrading the current strategies by hybridization and enhancing the accuracy.

3. METHODOLOGY

In order to expand the models BiLSTM -ANN and LSTM-ANN, we have got completed the six elements one after the other and linked them to every different in a series which includes records collection, preprocessing the raw facts, then separate the education and trying out statistics for the cause of education set of rules with validation, constitute the organized records into 300. dimensions of vectors for each term using the pre-educated fast text word embedding System[13].

3.1 Pre-processing

An initial step in text and sentiment classification is pre-processing. A significant amount of techniques is applied to data in order to reduce the noise of text, reduce dimensionality, and assist in the improvement of classification effectiveness. The most popular techniques include:

- Tokenization
- Remove Spelling
- Remove punctuation
- Remove Emotion
- Word Normalization
- Lemmatization

Tokenization :Refers to splitting up a sentence, phrase, or word into numerous smaller linguistic units named "Tokens". These tokens help to comprehend the NLP model and decipher the significance of the content by investigating the arrangement of the words. In natural language processing, the language needs to be analyzed and scrutinized under certain constraints and conditions. Tokenization breaks up a text to facilitate the whole process of analyzing a language in detail.

Remove Spelling :The collected text data was full of spelling blunders, and hence the raw data had to undergone In this dataset, most of the errors were type-error. Consequently, we made corrections to these words. This process was primarily done by a deep learning-based spell checker module which used the in word database as the training system. We pass every article and get a text without spelling errors.

Stop words removal: Stop words are the words being used in a language without coordinating meaningful information. In Bangla, there are a handful of stop words that do not denote any particular meaning rather than helping another phrase or words to make sense of. In order to train a model, it's very important to remove the stop words since stop words exist in a high quantity without providing any meaningful and unique sense.

Remove Punctuation l: Punctuation can play an important role when it comes to creating an emotional vibe to the expression. Apart from that, punctuation has no role when data has to be converted to a word embedding method. In our work, the word embedding method was incorporated, and for that reason, punctuations were to be removed beforehand.

Remove Emotion: The Internet has opened the horizon of globalization through social media. Emoticons play a vital role in expressing situational emotion while communicating with another person on social media. Nonetheless, the emoticons do not convey a message and do not contain any meaning themselves. Thereupon, Emoticons were removed from the corpus for further investigation.

Word normalization: Many words were not used in their standard form in our corpus. Some of the spellings were distorted, and some of them were in an informal format. We converted words into their original spelling automatically by python programming. This process is named word normalization, and it is a prerequisite for a well-developed NLP model [16].

Lemmatization: Lemmatization is the method of changing a word to its base structure [17]. The contrast between stemming and lemmatization will be, lemmatization considers the unique circumstance and stemming simply eliminates the last couple of characters, frequently prompting off base implications and spelling blunders. Lemmatization in phonetics is the way toward gathering the versatile types of a word recognized by word's lemma or word reference structure.

Data splitting: Training, validation, and testing are supposed to be the most important phase in the realm of machine learning. In order to train the models, the corpus was methodized in a way, so it becomes compatible with the computation process of the algorithms. Training allows the models to learn from their respective trials and error. Our dataset was split into three segments such as 70% and 10%, and 20%. 70% of our dataset was used for training the models, and the remaining 10% was kept for validation purposes, and the rest 20% is for the testing model. Validation helps the model to evaluate itself and grease the wheels for training repeatedly. After the models were trained, the testing data was used as the testing set for testing the model's performance.

3.2 Word Embedding Model

The word embedding model used here is fastText [18], a word embedding model developed by the Facebook research team. In traditional word embedding models, a vector with one target element is incorporated by attributing a numerical value of 1 and others 0. The length of the vector had to be huge and exhausting to analyze. Ideally, fastText is an extension to the word2vec model that represents a word by generating a vector summing up all the N- gram characters. The advantage of fastText is that it spawns a better word embedding scheme for the rare words and words out of the dictionary. In this work, a pre-trained fastText model with 300 dimensions has been utilized as a part of the word embedding tech.

3.4 Proposed Method

In this work, we integrated two neural network-based deep learning models for the content classification from Newspaper data. These are LSTM-ANN and BiLSTM-ANN. We also build the single LSTM and BiLSTM model so that the purpose can be analyzed based on the performance of the classification problem.These are LSTM-ANN and BiLSTM-ANN. We also build the single LSTM and BiLSTM model so that the purpose of the hybrid integration can be analyzed based on the performance of the classification problem. Embedding Layer is by and large used in the uses of Natural Language Processing, yet it can likewise be utilized with different assignments such as neural networks. In this investigation, the pre-trained word embeddings model fastText was incorporated. To manage text-based information, we need to change over it into numbers prior to implementing any machine learning algorithm. Moreover, all of the information esteems, including input values, are required to be in numerical form for neural networks.The embedding layer makes it possible to change over each word into a fixed-length vector of characterized size. The fixed length of word vectors expedites to address of words in an efficient manner and helps to decrease dimensions and calculation complexities In the embedding layer, the input dimension is the size of the vocabulary, and the output dimension is the length of the vector for each vocabulary. Here the accordant input dimension is in the embedding layer is 50000. The output dimension is 300, which is the dimension of the fastText pertained model. Input length refers to the length of input sequences when it is constant. The maximum input length is 200, which is the maximum sequence length of a text in our research. The L2 regularization penalty is computed as $loss = l2 * reduce\_(square(x))$. The embedding l2 regularization we have used here is 0.0005.Drop out layer is used to avoid overfitting. Since the results of a layer under dropout are arbitrarily subsampled, it lessens the limit of the network during training the models. To do that dropout layer randomly sets input units to 0 with a recurrence of the rate at each progression during preparing time. Input sources that are not set to 0 are scaled up by $1/(1 - rate)$ with the end goal that the total information sources are unaltered. After every LSTM and dense layer, we used a 0.04 dropout layer. Moreover, we utilized used dropout layer after embedding the layer.After the embedding layer was done, we passed the data to the BiLSTM or LSTM layer. We used two BiLSTM or LSTM layers for each epoch. The first layer contains 128 memory units and 64 for the next layer. The activation function used here is tanh, dropout is 0.02, and a recurrent dropout is 0.04. We have used both kernel regularizer and bias regularizer as 0.005 of l2 regularizer. Flatten is the activity of converting the pooled feature to a solitary column that is passed to a fully connected layer. In a way, it is discarding all the dimensions of a vector except one. A flatten procedure on a tensor reshapes the tensor to have a shape that is equivalent to the number of components contained in the tensor without. considering the batch dimension. If inputs are shaped without a feature axis, flattening adds an extra channel dimension and output shape (batch, 1). After passing both LSTM layers, we used the flatten layer to pass by the ANN.The dense layer is a neural network layer that is strongly connected, which means that any neuron in the dense layer receives feedback from all neurons in the previous layer. The dense layer is assumed to be the most widely used layer in models. Here we used three dense layers as the formation of ANN. The unit of the first layer is 128, and 64 for the next layer. For the first two dense layers, we used the regularizer function applied to the kernel weights matrix as kernel_regularizer and activation function 'Relu'. Regularizer function applied to the bias vectors a bias regularizer. Here bias regularizer = l2 regularizers of 0.001. Since we are solving a multiclass classification problem, we used the softmax activation function with one unit in the last dense layer.

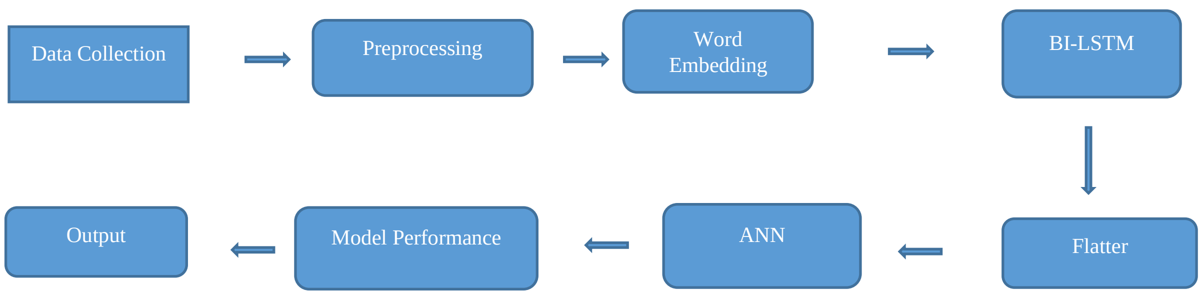

Fig -1 Methodology of Proposed Work

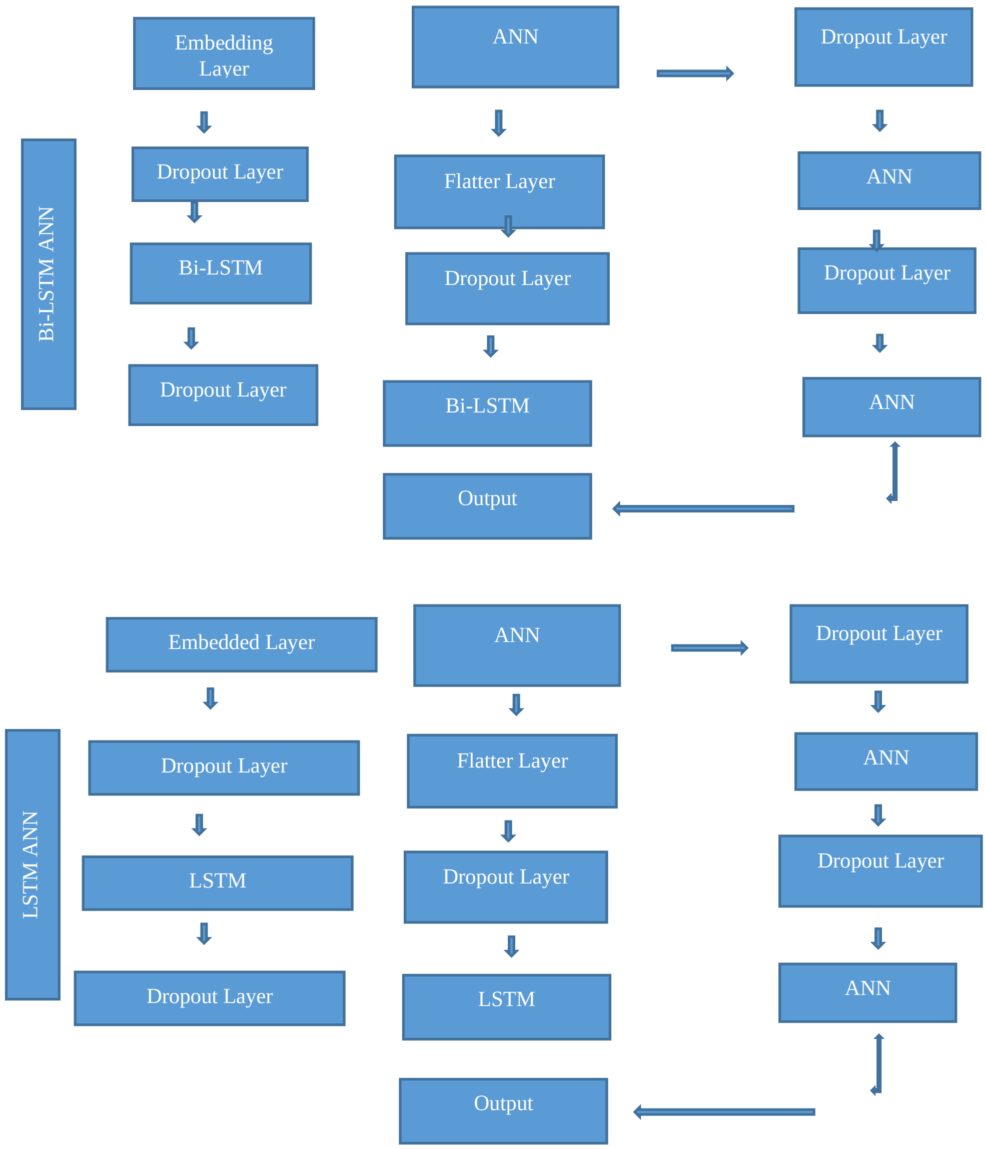

Fig 2 Bi-LSTM -ANN, LSTM_ANN

Tab1. Performance of Models

Models	Accuracy	F1 Score	Precision	Recall
BILSTM + ANN (Training)	0.96064	0.940 51	0.954615	0.944634
BiLSTM + ANN (Validation)	0.931891	0.90 0297	0.914304	0.893288
BiLSTM + ANN (Testing)	0.930101	0.90 1201	0.909435	0.890081
LSTM + ANN (Training)	0.946924	0.91 8872	0.929223	0.918764
LSTM + ANN (Validation)	0.915701	0.86 7182	0.878175	0.876227
LSTM + ANN (Testing)	0.909702	0.85 9781	0.871177	0.875202
BiLSTM	0.935001	0.90	0.906703	0.905775

(Training)		0021		
BiLSTM (Validation)	0.856177	0.77 1076	0.812058	0.764232
BiLSTM (Testing)	0.850101	0.78 1064	0.814071	0.769231
LSTM (Training)	0.901019	0.83 5804	0.842679	0.849088
LSTM(Valid ation)	0.800051	0.74 9745	0.902505	0.695793
LSTM(Valid ation)	0.801157	0.78 7425	0.922551	0.699353

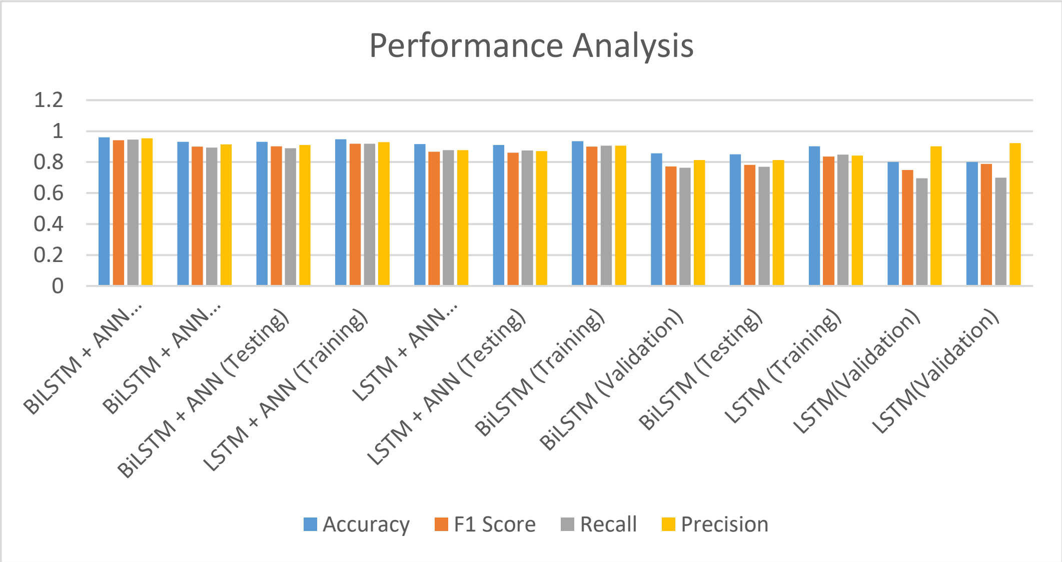

Fig 3 Perfomance Analysis

4. EXPERIMENT AND RESULT

This examination investigates how better or poorer the introduced hybrid LSTM-ANN and Bi-LSTM-ANN models perform in processing Bangla language to classify the contents from news articles. A robust experimental blueprint has been organized to get an insightful result from the proposed method. We arranged our experiment by developing the LSTM-ANN and BiLSTM -ANN models and further feeding our collected data to those models as depicted in the methodology section. We have split the data set into training (70%), validation (10%), testing (20%), and fed the corresponding section to trained and validated the models. Afterward, the test set was also exploited to test the performance of the models. In this section, we have studied how traditional LSTM and BiLSTM models perform compared to the proposed hybrid models classifying contents from Bangla text. To measure and analyze theperformance, we have utilized some statistical evaluation metrics. Besides, the important toolkit to develop the models and establish the workflow is also described. This section will assess the adequacy of our proposed strategy utilizing training sets and tuning significant hyper-parameters. The test dataset was utilized for assessing the models. We chose the hyperparameter by manual search and repeated the process until we reached an acceptable result. During this procedure, the hyperparameter is tweaked according to our observation and intuition based on previous researches. For every model, the assessment measurements considered were Accuracy, F1 Score (F1S), Recall (RS) and Precision (PS) . In addition, we also showed the ROC AUC curve to distinguish the performance of the models. Table 2 below depicts the exhibitions of order calculation in the resulting segment addressing the center discoveries of our study derived the techniques applied to accumulate and dissect data.

5.CONCLUSION

This paper shows that two deeply integrated models named BiLSTM-ANN and LSTM-ANN which can better perform the text content classification than the single development models like BiLSTM, LSTM and ANN. In order to validate the models with an experiment, we have used text articles from Bangla newspapers and blogs. To determine the performance of the proposed models, we have considered the six most relevant measurements from the performance metrics. Overall, the BiLSTM-ANN showed the best result in training, validation, and testing steps among all other models.

6. REFERENCES

[1] Sebåk M, Kacsuk Z. The Multiclass Classification of Newspaper Articles with Machine Learning: The Hybrid Binary Snowball Approach. Polit Anal 2021;29:236–49. https://doi.org/10.1017/pan.2020.27.

[2] Hidayatullah AF, Hakim AM, Sembada AA. Adult Content Classification on Indonesian Tweets using LSTM Neural Network. 2019 Int. Conf. Adv. Comput. Sci. Inf. Syst., IEEE; 2019, p. 235–40.

[3] Wang Y, Huang M, Zhu X, Zhao L. Attention-based LSTM for aspect-level sentiment classification. Proc. 2016 Conf. Empir. methods Nat. Lang. Process., 2016, p. 606–15.

[4] Nowak J, Taspinar A, Scherer R. LSTM recurrent neural networks for short text and sentiment classification. Int.

Conf. Artif. Intell. Soft Comput., Springer; 2017, p. 553–62.

[5] Kumar A, Singh JP, Dwivedi YK, Rana NP. A deep multi-modal neural network for informative Twitter content classification during emergencies. Ann Oper Res 2020:1–32.

[6] Yildirim O, Baloglu UB, Tan R-S, Ciaccio EJ, Acharya UR. A new approach for arrhythmia classification using deep coded features and LSTM networks. Comput Methods Programs Biomed 2019;176:121–33.

[7] Chen T, Xu R, He Y, Wang X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst Appl 2017;72:221–30.

[8] Sun Q, Amin M, Yan B, Martell C, Markman V, Bhasin A, et al. Transfer learning for bilingual content classification. Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. data Min., 2015, p. 2147–56.

[9] Dumais S, Chen H. Hierarchical classification of web content. Proc. 23rd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., 2000, p. 256–63.

[10] Lytvyn V, Burov Y, Kravets P, Vysotska V, Demchuk A, Berko A, et al. Methods and Models of Intellectual Processing of Texts for Building Ontologies of Software for Medical Terms Identification in Content Classification. IDDM, 2019, p. 354–68.

[11] Karim F, Majumdar S, Darabi H, Chen S. LSTM fully convolutional networks for time series classification. IEEE Access 2017;6:1662–9.

[12] Kowsher M, Islam Sanjid MZ, Das A, Ahmed M, Hossain Sarker MM. Machine Learning and Deep Learning based Information Extraction from Bangla Names. Procedia Comput Sci 2020; 178:224–33. https://doi.org/10.1016/j.procs.2020.11.024.

[13] Kowsher, Md. and Sobuj, Md. Shohanur Islam and Shahriar, Md. Fahim and Prottasha, Nusrat Jahan, Development of FasTtext Pre- Trained Model for Bangla NLP Research . Research on Computational Language,2021.

[14] Nawi NM, Atomi WH, Rehman MZ. The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks. Procedia Technol 2013;11:32–9. https://doi.org/10.1016/J.PROTCY.2013.12.159.

[15] Razavi AR, Gill H, Åhlfeldt H, Shahsavar N. A Data Pre-processing Method to Increase Efficiency and Accuracy in Data Mining. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2005;3581 LNAI:434–43. https://doi.org/10.1007/11527770_59.

[16] Toman M, Tesar R, Jezek K. Influence of Word Normalization on Text Classification n.d.

[17] Kowsher M, Tahabilder A, Hossain Sarker MM, Islam Sanjid MZ, Prottasha NJ. Lemmatization Algorithm Development for Bangla Natural Language Processing. 2020 Jt. 9th Int. Conf. Informatics, Electron. Vis. 2020 4th Int. Conf. Imaging, Vis. Pattern Recognition, ICIEV icIVPR 2020, 2020. https://doi.org/10.1109/ICIEVicIVPR48672.2020.9306652.

[18] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Trans Assoc Comput Linguist 2017;5:135–46. https://doi.org/10.1162/tacl_a_00051.

[19] Rehurek R, Rehurek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. Proc Lr 2010 Work NEW CHALLENGES NLP Fram 2010:45--50.