NOTAM Text Analysis and Classification Based on Attention Mechanism

Baigang Mi¹, Yi Fan¹*, Yu Sun²
¹School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China
²School of Foreign Studies, Northwestern Polytechnical University, Xi'an 710072, China
*email: fanyi@mail.nwpu.edu.cn

Abstract. As one of the important carriers, Notice to Air Men (NOTAM) provides key aeronautical security information during flight operation. However, a large number of NOTAM texts lack systematic utilization and organization. To improve the application of NOTAM texts, the Natural language processing can be used to classify and analyze NOTAM texts, and their effective processing attaches significance to monitor the air security. In this paper, a corpus set of NOTAM by web crawler technology has been built to generate raw data. Based on the aeronautical intelligence knowledge system, the corpus set is systematically pre-processed and analyzed, while the key information in NOTAM and corpus features are also extracted. Considering the high-dimensional and sparse of corpus set, the traditional bi-directional Recurrent Neural Network (Bi-RNN) architecture is improved by adding an attention layer which not only allows the Bi-RNN to capture hidden dependency to improve sparse features, but also allows the attention mechanism to weigh the key information base on their perceived importance to optimize text representation. Experiments show that the Attention-based Bi-RNN network outperforms the other four models, achieving a precision of 94.17% and an F1-Score of 93.66%. The model rapidly increases the utilization of text features meanwhile reduces the loss, which demonstrates it is a useful tool for the task of NOTAM texts classification and air security surveillance.

1. Introduction
As an important part of integrated aeronautical information series, NOTAM is a regulation or temporal change that pilots and relevant aviation personnel must understand and master the information, which is related to telecommunication regarding the establishment, prerequisites or trade in any aeronautical facility, service, system or hazard, the timed know-how of which is integral to personnel and structures involved with flight operations. Each NOTAM textual content contains, in addition to different information, a description of the records written in natural language, and every NOTAM textual content is assigned codes from predefined taxonomies. Complexity arises, on the one hand, from the want to categorize the textual content on the different hand, from the want to analyze and apprehend the reviews from a world factor of view. The universal purpose is to strengthen the equipment to assist categorize and analyze the data.

In recent years, with the rapid development of natural language technology, these techniques including text classification, text analysis and text mining have been widely and deeply applied in various domains. In general domain, Kim [1] pick up the essential data in sentences and extracted the nearby relevance of textual content by way of a couple of convolution kernels of exclusive sizes in...
Convolutional Neural Network (CNN). Zhang [2] additionally attempted the comparable CNN in textual content classification assignment and introduced character level to enhance the performance. But CNN is with extraordinarily bad overall performance in disposing longer sequence. Hao[3] and Cho[4] individually adopted Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) model to allow classification task. Besides, Yang [5] paid extra attention to archives classification via proposing a hierarchical attention network.

But the methods summarized above cannot entirely migrate into a specific domain, in agricultural domain, SVM (Support Vector Machine) model used to be carried out on agriculture dataset to consider classification project by means of Shi [6]. In scientific text, there are current mount of complicated features, a sensible technique was once taken with the aid of Hughes [7] to mechanically classify with a sentence degree in CNN. Sulea [8] amassed felony instances from the French Supreme Court and utilized he desktop getting to know strategies to predict the effects which made by means of judges. Rosa [9] targeted on navy chat or posts, which comprise gadgets of pastime and evaluated numerous present day textual content categorization and function determination methodologies on chat posts. Yang [10] proposed a novel methodology for producing conceivable counterfactual explanations, at the same time as evaluating the regularization advantages of adversarial coaching on language fashions in the area of finance. The researches above contain in medical, agriculture, military, regulation and economic domain, and in aeronautical talent facts that we lookup on is not often investigated. As for aviation incident reports, Tanguy [11] designed a number of NLP (Natural Language Processing) techniques to mine and control the information. Xu [12] described the textual content evaluation framework and used aviation datasets to put into effect NLP tasks. Alkhamisi [13] tried to discover a clever security supervision machine to administrate and predict the chance in aviation working through the ensemble of computer mastering algorithms. Rose [14] introduced a methodology for the evaluation of aviation protection narratives based totally on text-based bills of in-flight occasions and specific metadata parameters which accompany them.

The researches summarized above are characterized abundant corpus and great portability in general domain text analysis and classification task, However, owing to the situation that aeronautical information with spare corpus and high noise, the current text analysis and classification methods are not applicable to be directly used. To some extent, this restricts the intelligent analysis and development of aeronautical information intelligence.

In order to deal with the problems given above, we will build a NOTAM corpus set initially, then apply it into a NLP task. In the section one, through the web crawler technology, amount of NOTAM texts is fetched from the website. Based on the NOTAM rules, the key information in NOTAM is extracted to make up the corpus set. In the section two, we apply the corpus to achieve a text classification task, which is helpful to aviation bodies to address innumerable NOTAM texts. To improve the performance of text classification, we adjust and optimize some key parameters in the model.

2. Construction of corpus set
Aeronautical corpus performs an quintessential position in natural language processing task, alas there is little corpus in aviation enterprise to supply the researchers to mine the information further. National General Aviation Flight Information Database presently homes over 700,000 h of per-second time collection flight statistics recorder readings generated via over 400,000 flights from eight fleet of plane and over one hundred ninety taking part personal individuals.[15] A public database of cosmic radiation measurements at aviation blanketed greater than 4500 flights with greater than a hundred thirty zero measurements.[16] Flight extend prediction machine constructed a dataset for the proposed scheme, computerized structured surveillance-broadcast (ADS-B) messages are received, pre-processed, and built-in with different records such as climate condition, flight schedule, and airport information.[17] Thus, we will build a corpus set about NOTAM to solve the problems.
2.1. Corpus collection
NOTAM covers abundant contents and change swiftly so they are transmitted by the aeronautical fixed telecommunication network (AFTN). In order to satisfy transmitting, NOTAM must be issued following a specific format and using the specific characters. Because of this, we collect raw NOTAM data from Federal Aviation Administration through web crawler technology. The raw data is cleaned and filtered by regular expressions to remove useless information and extract 98088 raw NOTAM texts. Table 1 shows a typical format of raw NOTAM that crawler from web.

Table 1 Example of raw NOTAM

| Item       | Content                                                                 |
|------------|-------------------------------------------------------------------------|
| ID         | U1973/21 NOTAMN                                                         |
| Qualifiers | Q) ZPKM/QNMCT/IV/BO/AE/000/999/3035N10357E025                         |
| Time and Location | A) ZUUU B) 2106201930 C) 2106220100                             |
| Text       | E) CHONGZHOU VOR/DME 'CZH' 114.5MHZ/CH92X U/S DUE TO FLTCK.          |

As is shown in Table 1, a complete NOTAM text is including ID number, qualifiers, time and location information, and text. Especially, each piece of content express information referring to aeronautical intelligence. As is shown in figure 1, the basic analysis of a NOTAM text is clearly implied.

Fig. 1 Basic Information of NOTAM

2.2. Text analysis

2.2.1. Fundamental information Fundamental information is referring to the ID, qualifiers, and time and location. This information must adhere to a fixed format, which means it is easy to extract. Meanwhile, they extremely rely on the aeronautical intelligence rules. We will use regular expression to summarize them into a corpus set. And a part of hidden features are mining, including the length, the duration time, and importance. The manual work also is taken to check whether the information is accurate. An example of fundamental information has been processing is shown in Table 2. Also scope information statistical figure is shown in Figure 2.
2.2.2. Main body text. The main body text cannot be directly input into any tasks due to the features of disorder, the texts should be initially disposed by the means that NLP usually adopts in general domains. The punctuations and the stops words are dealt by loading vocabulary, cleaning and filtering the special characters that influenced features extracting, and segment words by the word segmentation toolkit. As for the distinctiveness of NOATAM corpus, our work is involved as follows:

1) Abbreviation analysis: totally 996 pairs of abbreviation are replaced according to the list of abbreviation provided by International Civil Aviation Organization (ICAO).

2) Digital reservation: although data information often is ignored in general domain, data information is expressed radio frequency, runway number, distance, azimuth, flight level and other information. As an essential feature, this information can help to improve the accuracy of text classification task. The comparison of text before and after analysis is presented in Table 3:

| ID  | New Theme | Type  | Min_flt | Radius | Effective time |
|-----|-----------|-------|---------|--------|----------------|
| U3171 | TRUE landing area IV | 0 | 5 | 2021/8/28/18:00 |
| Year | FIR | Situation Purpose Max_flt Aerodrome Failure time |
| 2021 | ZPKM | Limitation NBO 999 ZUUU 2021/8/28/22:00 |
| Length | Scope Permanent Longitude Latitude Duration Importance |
| 35 | A | FALSE 103.56E 30.34N 4 False |

According to the weights of TF-IDF, the extracted concepts are sorted, and the 100 words thermodynamic diagram with the highest weight of main body text is shown in Figure 3.
3. Application of corpus set
As for NOTAM managing procedures, it is difficult for aeronautical intelligence personnel to dispose massive NOTAM texts with high frequency and accuracy. But each NOTAM text is not equally important, how to recognize the more important NOTAM text is affecting to avoid the accidents and impoverish the security of aviation operating. With the information extracting, a text classification task based on deep learning network is implemented to distinguish which NOTAM should be disposed with a high-priority rating. In this task, it will be divided into three sections: data calling, model building and experiment. Thus, we call the theme and main body information in corpus set to accomplish the task. Observing the information, those includes 43133 texts and the distribution of them is shown in Table 4.

| Tab. 4 Distribution of NOTAM texts in text classification task |
|---------------------------------------------------------------|
| category                                      | code | amount |
| lighting facilities                                    | L    | 1132   |
| movement and landing area                             | M    | 11219  |
| facilities and services                               | F    | 5037   |
| communication and radar facilities                    | C    | 1082   |
| ILS and MLS                                           | I    | 1588   |
| terminal &enroute navigation facilities               | N    | 2409   |
| GNSS (Global Navigation Satellite System)             | G    | 152    |
| service                                                |      |        |
| airspace organization                                  | A    | 2695   |
| air traffic and VOLMET services                       | S    | 1554   |
| air traffic procedures                                 | P    | 3975   |
| airspace restrictions                                  | R    | 4289   |
| warnings                                               | W    | 5273   |
| other information                                      | O    | 2310   |
| checklist                                              | K    | 418    |

3.1. Model building
Attention-based RNN model not only allows the Bi-RNN network to capture hidden dependency to improve sparse features, but also allows the attention mechanism to weigh the key information base on their perceived importance to optimize text representation. The architecture of attention-based Bi-RNN is shown in Figure 4.
Input layer contains $n$ words expressed as a sequence of $W_1, W_2, \ldots, W_n$. Word embedding layer transmits pre-training word vectors into a fixed-dimension word embedding matrix, also input them into Bi-RNN model. RNN neural network mines word context semantic features and captures some word-to-word dependencies, but a part of hidden dependencies has lost. Then introducing Bi-RNN, with the propagation of forward and backward RNNs alternately, context information and capturing hidden context dependencies are tightly linked to optimize the sparse semantic features of short text. Then, a representation will generate and combine with the RNN. Finally, the classifier, softmax layer, predicted the results by the representation feeding from attention layer.

3.1.1. **Bi-RNN model** LSTM recurrent neural network can be broken into three parts: input gate, forget gate, and output gate. Each gate is corresponding with weight matrix: input gate is corresponding with $W_{xi}, W_{hi}, W_{ci}, b_i$, forget gate is corresponding with $W_{xf}, W_{hf}, W_{cf}, b_f$, and output gate is corresponding with $W_{xo}, W_{ho}, W_{co}, b_o$. The LSTM cell state input $x_t$ controlled by a gating mechanism, previous layer generates a state $h_{t-1}$ and updates gating by equations (1)(2), then updates inner state $c_t$ by equation (3), last updates hidden state $h_t$ by equations (4)(5). The function for gates and cell states are computed as:

\[
i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)
\]
\[
f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)
\]
\[
o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o)
\]
\[
g_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + W_{cc}c_{t-1} + b_c)
\]
\[
c_t = i_t g_t + f_t c_{t-1}
\]
\[
h_t = o_t \tanh(c_t)
\]

The development of LSTM network has led to many network variants, such as GRU network. GRU and LSTM have similar performance. Both of these models can effectively handle gradient
disappearance or explosion, and capture long-range dependencies. They have simple structures, high computational efficiency, few network parameters and small storage spaces. GRU recurrent neural network can be broken into two parts: update gate, and reset gate. Each gate is corresponding with weight matrix: update gate is corresponding with \( W_{zu}, W_{zh}, b_r \), and reset gate is corresponding with \( W_{zx}, W_{zh}, b_z \). GRU cell state input \( x_t \) controlled by a gating mechanism, previous layer generates a state \( h_{t-1} \) and updates gate output \( r_t, z_t \), by equations (7)(8), then updates candidate state \( h'_t \) by equation (9), last updates hidden state \( h_t \) by equations (10). The function for gates and cell states are computed as:

\[
\begin{align*}
    r_t &= \sigma (W_{zu} x_t + W_{zh} h_{t-1} + b_r) \\
    z_t &= \sigma (W_{zx} x_t + W_{zh} h_{t-1} + b_z) \\
    h'_t &= \tanh (W_{h0} x_t + W_{h1} r_t h_{t-1} + b_h) \\
    h_t &= (1 - z_t) h_{t-1} + z_t h'_t
\end{align*}
\]

Traditional RNN given above capacity that they solely study the hidden facts from preceding time step. Bi-RNN layer consists of two sub-layers: one shifting ahead in time steps and one transferring backwards in time steps. \( \tilde{h} \) and \( \tilde{h} \) are the hidden states of the forward and backward traversals \( w_t \). To compute the hidden state, we perform an element-wise sum of the hidden states computed from both its sublayers by equations (6)(10).

\[
M = \tanh (H) \\
\alpha = \text{softmax}(W^{(3\prime)} M) \\
r = H \alpha^T \\
\tilde{h}^* = \tanh (r)
\]

Where \( M \in \mathbb{R}^{d_r \times n} \), \( H \in \mathbb{R}^{d_r \times n} \), \( d_r \) is embedding dimension, \( W^{(3)} \) is parameter vector from training, \( W^{(3\prime)} \) is parameter vector after transposition. The dimension of \( W, \alpha, r \) is individually \( d_r, n, d_r \). The representation of the final sentence \( h^* \) is active by the function \( \tanh \).

3.1.2. Classification layer. When the representation has been acquired from attention layer, it will convey to the softmax layer and the classification results which will be predicted. Then the softmax classifier transfers the output \( y_t \) into a probability \( p(k_i \mid D, \theta) \), and assigning the text class with the highest probability:

\[
p(k_i \mid D, \theta) = \frac{\exp(y_i)}{\sum_{j=1}^{n} \exp(y_j)}
\]

\[
\theta = \{E, W_{si}, W_{hi}, W_{ei}, b_i, W_{se}, W_{fe}, b_f, W_{so}, W_{ho}, W_{co}, b_o\}
\]

\[
E = [e(w_1), e(w_2), \ldots, e(w_n)]
\]
Where $D$ is input document, $\theta$ is parameter learning from training model, $k_i$ is classification marker.

3.2. Experiment

3.2.1. Design and performance of experiment The experimental data is extracted from corpus set totally 43133 pieces, and then divided them into train set, validation set and test set according to the ratio of 8:1:1. The specific experimental parameters are as follows: word vector dimension embedding _Size = 300, the size of neurons in the whole connective layer is hidden_ Dim = 128, learning rate= 0.001, batch size = 128, random inactivation of 50% of the nerve unit drop_ Out = 0.5, epochs = 10.

In the hypothesis of assuming experiment, we take cross-entropy loss, precision, accuracy, recall, and F1-Score as indicators of comprehensive evaluation. Table5. lists the index values of attention-based Bi-LSTM and Bi-GRU for NOTAM classification, including accuracy, recall and F1score. According to the three assessment values, each index performed well. Overall, Bi-LSTM model outperform than Bi-GRU model, F1 score individually has reached 93.66% and 93.04%.

**Tab. 5 Multi-indicator Effects of NOTAM Classification Tasks**

| Category | Precision(%) | Recall(%) | F1-Score(%) |
|----------|--------------|-----------|-------------|
|          | Bi-LSTM      | Bi-GRU    | Bi-LSTM     | Bi-GRU    | Bi-LSTM | Bi-GRU |
| L        | 85.37        | 85.25     | 92.11       | 91.23     | 88.61   | 88.14  |
| M        | 96.89        | 97.02     | 97.06       | 95.81     | 96.97   | 96.41  |
| F        | 90.32        | 89.11     | 90.67       | 90.87     | 90.50   | 89.98  |
| C        | 89.72        | 88.50     | 88.07       | 91.74     | 88.89   | 90.09  |
| I        | 97.44        | 95.15     | 95.60       | 98.74     | 96.51   | 96.91  |
| N        | 97.08        | 97.90     | 96.88       | 96.68     | 96.88   | 97.29  |
| G        | 100.00       | 81.25     | 81.25       | 81.25     | 89.66   | 81.25  |
| A        | 90.28        | 93.01     | 96.30       | 93.70     | 93.19   | 93.36  |
| S        | 88.27        | 83.93     | 91.67       | 90.38     | 89.94   | 87.04  |
| P        | 98.67        | 96.43     | 92.96       | 94.97     | 95.73   | 95.70  |
| R        | 97.64        | 95.82     | 96.50       | 96.27     | 97.07   | 96.05  |
| W        | 98.47        | 98.43     | 97.35       | 95.27     | 97.90   | 96.82  |
| O        | 88.19        | 94.27     | 90.48       | 92.64     | 89.32   | 93.45  |
| K        | 100.00       | 100.00    | 100.00      | 100.00    | 100.00  | 100.00 |
| Average  | 94.17        | 92.58     | 93.35       | 93.54     | 93.66   | 93.04  |

It can be seen from Table5. that compared with other texts, the K-class about the check list text has significant numerical features and fewer data samples, therefore the classification results all reaches 100%. It is noteworthy that the main reason for the effectiveness of class G GNSS texts in Bi-GRU model is that the number of texts is small, and the sensitivity of Bi-GRU model to sample size is different from that of the Bi LSTM model; The precision of L-class about lighting facilities text and S-class about meteorological information, texts are lower, and they have many similar features with M-class operation landing area text and P-class air traffic rule text. Furthermore, these four types of text are short in length, but there are large amounts of data in M-class and P-class text. This suggests that the amount of data speculated affects the effect of model adjustment to some extent. Therefore, according to the unbalanced characteristics of data volume, based on the above model, this paper further explores the impact of balanced and unbalanced data volume on the classification of the model. At the same time, the maximum text length of the model is set to 32, 64 and 128 characters for comparison. The evaluation index mainly adopts F1 score, and the results are displayed on Table 6.
Tab. 6 Comparison of F1-Score between Four Categories Text

| Category | 128          | 64          | 32          |
|----------|--------------|-------------|-------------|
|          | balanced     | unbalanced  | balanced    | unbalanced  | balanced    | unbalanced  |
| L        | 88.14        | 88.31       | 85.48       | 85.45       | 84.68       | 84.26       |
| S        | 88.27        | 85.29       | 86.98       | 85.19       | 85.9        | 84.98       |
| M        | 96.42        | 96.93       | 96.4        | 96.54       | 95.18       | 96.27       |
| P        | 96.73        | 96.82       | 94.35       | 95.35       | 95.52       | 94.97       |

Table 6 shows that F1 score of most categories is better than that of unbalanced data set in case of balanced data set. In the case of balanced case, F1 score of L-class and S-class text is higher than that of unbalanced case, while F1 score of M-class and P-class text is mostly lower than balanced situation in unbalanced case. Considering the difference between the number of L-class, S-class, M-class and P-class text. It is suggested that the data quantity has a certain influence on the classification effect of the model, which is consistent with the above Table 5. conclusion. In addition, it can be seen from Table 6. that with the increase of the maximum text length, the F1 score of model classification increases.

![Classification F1-Score of NOTAM Text in Different Max_Len](image)

3.2.2. Comparison of different model. In order to verify the effectiveness of the model, four groups of controlled experiments are designed, which are conducted on fasttext, RNN, CNN and Bi-RNN models respectively. During the experiment, the same basic super parameters are set for each model: random inactivation rate is 0.5, epoch is 10, and the maximum text length is 128. Table 6 shows the comparison of different deep learning models in accuracy, recall and F1 score. By comparing each model, the F1 score of the recurrent neural network model combined with attention mechanism is significantly higher than other models, which shows that the model can enhance the feature extraction and transmission, reduce the loss of features, and improve the classification effect.

Tab. 7 Comparison of NOTAM Classification between Different Neural Networks

| Model       | Precision(%) | Recall(%) | F1-Score(%) |
|-------------|--------------|-----------|-------------|
| FasText     | 88.91        | 84.50     | 86.44       |
| RNN         | 91.51        | 90.75     | 90.98       |
| CNN         | 91.14        | 92.15     | 91.60       |
| BiRNN       | 92.67        | 92.09     | 92.23       |
| ATT-BiLSTM  | 94.17        | 93.34     | **93.66**   |
| ATT-BiGRU   | 92.58        | 93.54     | **93.03**   |

With different text length, the classification accuracy of the model on balanced and unbalanced data sets is shown in Figure 5. It can be seen that with the increase of the maximum text length, the accuracy of model classification increases in both balanced and unbalanced data sets. On the whole, the prediction result of the model in the balanced situation is better than that in the unbalanced situation.
Figure 6. and Figure 7. show the accuracy and loss of different models on the validation data set. The attention-based Bi-LSTM and attention-based Bi-GRU models will not have a big shock in the training process, and will converge quickly and keep the highest in the subsequent training process, which further reflects the advantages of the model.

![Graph 6](image1.png) ![Graph 7](image2.png)

**Fig. 6** Different Neural Networks of Validation Accuracy Trend Graph  
**Fig. 7** Different Neural Networks of Validation Loss Trend Graph

4. Conclusion

In this paper, we aim at the corpus building and its application of NOTAM texts. By integrating and analyzing the corpus according to the characteristics of the domain, and we develop the deep learning method to build a Bi-RNN model combined with attention mechanism to complete the text classification task. This model can capture the dependency information hidden in context, and pay more attention to the key information of text, so as to improve the accuracy of text classification. The main conclusions are as follows:

1. A NOTAM corpus set is established by web crawler technology and text analysis theory. Qualitatively, this corpus is entirely following the basic rules of aeronautical intelligence and can be reused or shared with other aviation NLP task. Quantitatively, it covers 98088 NOTAM texts with finely processing and the data features in it can be expressed visually.

2. The attention-based Bi-RNN neural network is used to build the automatic text classification model of NOTAM, which can realize the automatic text classification in 14 categories quickly and accurately, and avoid the problems of time-consuming and poor classification effect of the shallow learning model. The accuracy and F1 score of the model reach 94.17% and 93.66%, which is similar to the other four text classification models (fasttext model, CNN model RNN model and Bi-RNN model);

3. By comparing the text classification results of different categories, we can see that the categories with large amount of data are more sensitive to the attention Bi-RNN model. By adjusting the maximum length and the balance of text, the problem of insufficient data can be alleviated to a certain extent, On the whole, with the increase of the maximum length, F1 score increases correspondingly, and the classification results of balanced data sets are better than those of unbalanced data sets.

There is a large amount of information in NOTAM, and the analysis and classification methods vary with the content and users. This paper explores only one way to process texts based on deep learning. In the next step, we need to consider other NLP systems of NOTAM, and start from the perspective of reporting the status of transactions and emergencies, to realize the multi angle and all-
round processing of notations; In addition, the accuracy of the model can also be improved by adjusting the parameters of feature fusion, so as to achieve the method of accurate detection of aviation safety.

Acknowledgment
The authors would like to acknowledge the support of the Soft science program of Shaanxi Province (Grant No. 2021KRM187).

References
[1] Kim Y 2019 Convolutional neural networks for sentence classification. arXiv 2014. arXiv preprint arXiv:1408.5882.
[2] Zhang X, Zhao J and LeCun Y 2015 Character-level convolutional networks for text classification Advances in neural information processing systems 28 pp 649-657
[3] Hao Y, Zhang Y, Liu K, He S, Liu Z, Wu H and Zhao J 2017 An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics vol 1 pp 221-231
[4] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation arXiv preprint arXiv:1406.1078.
[5] Yang Z, Yang D, Dyer C, He X, Smola A and Hovy E 2016 Hierarchical attention networks for document classification Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies pp 1480-1489
[6] Shi L, Duan Q, Ma X and Weng M 2011 The research of support vector machine in agricultural data classification International Conference on Computer and Computing Technologies in Agriculture (Springer, Berlin, Heidelberg) pp 265-269
[7] Hughes M, Li I, Kotoulas S and Suzumura T 2017 Medical text classification using convolutional neural networks Informatics for Health Connected Citizen-Led Wellness and Population Health pp 246-250
[8] Sulea O M, Zampieri M, Malmasi S, Vela M, Dinu L P and Van Genabith J 2017 Exploring the use of text classification in the legal domain arXiv preprint arXiv:1710.09306.
[9] Rosa K D and Ellen J 2009 December Text classification methodologies applied to micro-text in military chat International Conference on Machine Learning and Applications (IEEE) pp 710-714
[10] Yang L, Kenny E M, Ng T L J, Yang Y, Smyth B and Dong R 2020 Generating plausible counterfactual explanations for deep transformers in financial text classification arXiv preprint arXiv:2010.12512.
[11] Tanguy L, Tulechki N, Urieli A, Hermann E and Raynal C 2016 Natural language processing for aviation safety reports: From classification to interactive analysis Computers in Industry 78 pp 80-95
[12] Xu B and Kumar S 2015 A text mining classification framework and its experiments using aviation datasets Proceedings of the Workshop of Information Technology and Systems Dallas
[13] Alkhamisi, A. O and Mehmoood, R 2020, March An ensemble machine and deep learning model for risk prediction in aviation systems In 2020 6th Conference on Data Science and Machine Learning Applications (CDMA) (IEEE) pp 54-59
[14] Rose R L, Puranik T G and Mavris D N 2020 Natural Language Processing Based Method for Clustering and Analysis of Aviation Safety Narratives Aerospace 7(10) p 143
[15] Karboviak K, Cluchar S, Desell T, Dusenbury M, Hedrick W, Higgins J, et al. 2018 Classifying aircraft approach type in the national general aviation flight information database International Conference on Computational Science (Springer, Cham) pp 456-469
[16] Kákona M, Kyselová D, Ambrožová I, Kubančák J, Štipán V, Langer R, et al. 2019 CR10—A public database of cosmic radiation measurements at aviation altitudes of about 10 Km Radiation protection dosimetry 186(2-3) pp 224-228

[17] Gui G, Liu F, Sun J, Yang J, Zhou Z and Zhao D 2019 Flight delay prediction based on aviation big data and machine learning IEEE Transactions on Vehicular Technology 69(1) pp 140-150

[18] Sutskever I, Vinyals O and Le Q V 2014 Sequence to sequence learning with neural networks Advances in neural information processing systems pp 3104-3112

[19] Bahdanau D, Chorowski J, Serdyuk D, Brakel P and Bengio Y 2016 End-to-end attention-based large vocabulary speech recognition 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP) (IEEE) pp 4945-4949