Tech Science Press

Sammon Quadratic Recurrent Multilayer Deep Classifier for Legal Document Analytics

Divya Mohan* and Latha Ravindran Nair

School of Engineering, Cochin University of Science and Technology, Kerala, India
*Corresponding Author: Divya Mohan. Email: divyamohancusat@yahoo.com

Abstract: In recent years, machine learning algorithms and in particular deep learning has shown promising results when used in the field of legal domain. The legal field is strongly affected by the problem of information overload, due to the large amount of legal material stored in textual form. Legal text processing is essential in the legal domain to analyze the texts of the court events to automatically predict smart decisions. With an increasing number of digitally available documents, legal text processing is essential to analyze documents which helps to automate various legal domain tasks. Legal document classification is a valuable tool in legal services for enhancing the quality and efficiency of legal document review. In this paper, we propose Sammon Keyword Mapping-based Quadratic Discriminant Recurrent Multilayer Perceptive Deep Neural Classifier (SKM-QDRMPDNC), a system that applies deep neural methods to the problem of legal document classification. The SKM-QDRMPDNC technique consists of many layers to perform the keyword extraction and classification. First, the set of legal documents are collected from the dataset. Then the keyword extraction is performed using Sammon Mapping technique based on the distance measure. With the extracted features, Quadratic Discriminant analysis is applied to perform the document classification based on the likelihood ratio test. Finally, the classified legal documents are obtained at the output layer. This process is repeated until minimum error is attained. The experimental assessment is carried out using various performance metrics such as accuracy, precision, recall, F-measure, and computational time based on several legal documents collected from the dataset. The observed results validated that the proposed SKM-QDRMPDNC technique provides improved performance in terms of achieving higher accuracy, precision, recall, and F-measure with minimum computation time when compared to existing methods.

Keywords: Legal document data analytics; recurrent multilayer perceptive; deep neural network; sammon mapping; quadratic discriminant analysis; likelihood ratio test

1 Introduction

In common law legal systems, judges make decisions for issues between parties by reference to previous decisions that consider similar factual situations. A new decision frequently cites previous relevant decisions and such decisions (citations) may form rich citation networks. Such labeling of cases is important to the process of determining whether a case is proper law. These citations help lawyers and students to easily look up case files and documents they are searching in a pile of other documents. These labels serve as a matter of convenience in citation indices enabling lawyers to prioritize decisions to examine and to understand the current state of the law. These also helps in predicting the results of legal cases and hence in making the legal decision-making process faster. Legal text mining aims to automatically evaluate the texts in the legal domain. Efficient document classification techniques are applied for Legal text mining. However, the conventional machine learning-based classification techniques are often inefficient for achieving higher accuracy.

In recent years deep learning techniques has shown promising results in the field of legal domain. Neural networks such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been used for various legal domain tasks such as information retrieval, document classification, document summarization, etc.

A Joint Bidirectional Label Attention Conditional Network (JBLACN) model was introduced in [1] to classify the court record documents based on the extracted evidence information with higher precision and recall. But the accuracy of the court record documents classification was not improved while considering the more documents. A novel hierarchically nested attention structure model was introduced in [2] for predicting the classification of judgment documents. But, the designed method failed to use the more complex legal judgment cases for accurate document classification.

An active learning method was developed in [3] for document classification. But the deep learning-based classification was not performed to further increase the overall accuracy. In [4], data mining techniques were introduced for smart legal systems but the automatic keyword extraction from legal data was not performed. A controversial issues classification algorithm was designed in [5] to improve the accuracy. But the time consumption of the classification was not reduced.

A long–short-term memory (LSTM) based recurrent neural networks were introduced in [6] for document classification. However, the model failed to use the efficient classification approach for the capability of training the model. A bidirectional Recurrent Neural Network was introduced in [7] to improve the text classification accuracy. But the performance of time consumption was not minimized.

A Graph LSTM (Long Short-Term Memory) was introduced in [8] based on domain knowledge extraction for Chinese legal document categorization. Though the model reduces the time complexity, the classification was not performed through semantic analysis. The machine learning techniques were developed in [9] for categorizing the text documents to predict the ideological way of judgment from the associated text. However, the performance of accuracy and precision of different methods was not analyzed.

An ontology-driven knowledge block summarization method was introduced in [10] for categorizing the judgment documents based on the similarity. However, it failed to explore the novel techniques to investigate for the most similar case judgments.

A local convolutional feature aggregation technique was introduced in [11] to execute document categorization. The designed technique failed to get better parameters to improve the performance. A semi-supervised approach was presented in [12] to improve the accuracy and precision of the document categorization. But the time consumption of the designed approach was not minimized.

A Multichannel Attentive Neural Network (MANN) was introduced in [13] for judgment documents classification. But the designed network failed to provide better performance. A neural network language modeling (NNLM) approach was designed in [14] to analyze and recognize the judgment documents. However, the quantitative evaluations of the system, incorporating legal expert evaluation were not performed. In [15], different machine learning classification algorithms were developed. But it is difficult to automatically extract the features from the text of judgment while enabling the collection of a large number of structured documents.

A deep learning-based robust reconstruction model was introduced in [16] to facilitate faster strip-shredded documents classification. However, the designed deep learning model failed to extract text information from the reconstructed documents. A target center-based LSTM model was introduced in [17] for performing the task of keywords extraction. But, it failed to focus on finding the semantic relations between different words.

A Topical Position Rank model was introduced in [18] for keyphrase extraction and key phrase classification. But the deep learning model was not implemented to enhance the accuracy of classification. An improved TextRank model was introduced in [19] for increasing accuracy and minimizing the complexity of document classification. However, the model failed to combine more heterogeneous information for keywords rank value estimation. A Maximum Entropy Partitioning (MEP) method was introduced in [20] for extracting the keywords to perform document analysis. But the method was not efficient to decrease the time complexity.

A deep learning architecture was proposed in [21] that adopts domain-specific pre-training and a label-attention mechanism for multi-label document classification. A deep neural network model was proposed in [22] for coreference resolution in court record documents. An efficient and appropriate safety testing tool was introduced in [23] to identify the source of printed documents is an important task in the meantime.

The neural model with the Simulated Annealing (SA) a classic global optimization algorithm with low computational complexity was developed in [24] to reduce the daily training time, providing a more friendly graphic visualization of documents in high dimensions, supporting the judicial decision process. An exhaustive and unified repository of judgments documents, called ECHR-OD was presented in [25] based on the European Court of Human Rights.

Sl. no	Title	Method	Proposal	Merits & Demerits
1	Identify trademark legal case precedents - Using machine learning to enable a semantic analysis of judgments	Neural network language modeling (NNLM) approach	A neural network language modeling (NNLM) approach was designed in [14] to analyze and recognize the judgment documents	TM litigation trends as better protect for brand equity. The quantitative evaluations of the system, incorporating legal expert evaluation were not performed

(Continued)

CMC, 2022, vol.72, no.2

Continued

Sl. no	Title	Method	Proposal	Merits & Demerits
2	Self-supervised Deep Reconstruction of Mixed Strip-shredded Text Documents	Deep learning-based robust reconstruction model	A deep learning-based robust reconstruction model was introduced in [16] to facilitate faster strip-shredded documents classification	Small and local samples are more effective for learning the compatibility between shreds. The designed deep learning model failed to extract text information from the reconstructed documents.
3	Keywords Extraction with Deep Neural Network Model	Target center-based LSTM model	A target center-based LSTM model was introduced in [17] for performing the task of keywords extraction	Deep learning algorithm improves the performance of many tasks Target center-based LSTM is failed to focus on finding the semantic relations between different words
4	TOP-Rank: A TopicalPostionRank for Extraction and Classification of Keyphrases in Text	Topical Position Rank model	A Topical Position Rank model was introduced in [18] for keyphrase extraction and key phrase classification	Ranking SVM makes proper use of tagging information and achieves significant performance enhancement in keyphrase (tag) extraction. Deep learning model was not implemented to enhance the accuracy of classification.
5	A patent keywords extraction method using TextRank model with prior public knowledge	Improved TextRank model	An improved TextRank model was introduced in [19] for increasing accuracy and minimizing the complexity of document classification	Patent keywords can be extracted by finding top-k node words with higher node rank values TextRank model failed to combine more heterogeneous information for keywords rank value estimation

The major contribution of the proposed SKM-QDRMPDNC on the contrary to the conventional method is summarized as given below,

- To increase the accuracy of legal document classification, a novel SKM-QDRMPDNC technique is introduced based on two different processes namely keyword extraction, and classification.

- To extract keywords from the documents based on the frequency score, novelty of Sammon Mapping technique is applied in SKM-QDRMPDNC technique. Then distance measure is performed among frequency score and threshold value. Depending on the distance measures machine learning technique transforms keywords from high-dimensional space into lower dimensionality. This helps to decrease computational time.
- To perform the document classification based on the likelihood ratio test, SKM-QDRMPDNC technique uses novelty of Quadratic Discriminant analysis in second hidden layer of deep learning. The likelihood ratio is used to classify the documents with higher accuracy.
- Finally, experimental evaluation is performed to estimate the quantitative analysis of the proposed SKM-QDRMPDNC technique with the existing algorithms and different performance metrics like accuracy, precision, recall and F-score have been used for this purpose.

The remainder of this paper is organized into different sections. In Section 2, the proposed SKM-QDRMPDNC technique is described in detail. Section 3 shows the experimental analysis. Quantitative discussion of the proposed technique and existing methods are presented in Section 4. Finally, the conclusion is presented in Section 5.

2 Sammon Keyword Mapping Based Quadratic Discriminant Recurrent Multilayer Perceptive Deep Neural Classification

A novel SKM-QDRMPDNC technique is introduced for improving the legal document analysis. For large number of legal texts, significant keywords extraction is a challenging task. Therefore, the proposed SKM-QDRMPDNC technique uses a dimensionality reduction technique to extract the significant keywords from the legal documents to minimize the time complexity of document classification.

Fig. 1 demonstrates the structural design of the proposed SKM-QDRMPDNC technique consisting of two major processes for improving accuracy. Initially, the set of legal documents are gathered from the dataset. After that, the important keywords are extracted for accurate classification with minimum time. Secondly, the classification is done with the help of a correlation coefficient. The proposed SKM-QDRMPDNC technique uses the recurrent multilayer perceptive deep neural classification where the structure uses many layers of nodes to obtain high-level functions from input information.
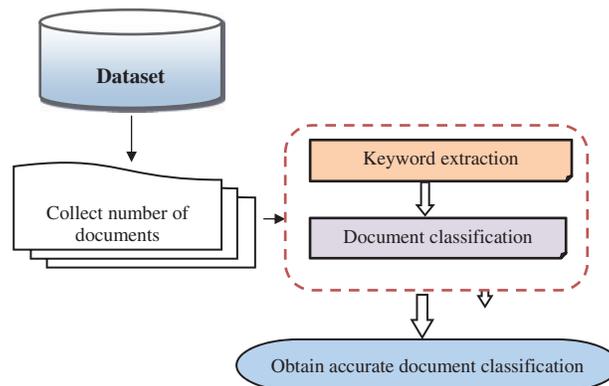

Figure 1: Architecture of the SKM-QDRMPDNC technique

SKM-QDRMPDNC technique helps a model to increase accuracy. It also allows a model to take a set of inputs and provide an output. Each of these layers is feed-forward except for the last layer (*i.e.*, output), which has feedback connections to the hidden layer.

Fig. 2 illustrates the structural design of recurrent multilayer perceptive deep neural classification. A recurrent deep neural classification is an artificial neural learning network that consists of three layers of nodes - an input layer, a hidden layer, and an output layer. The input layer receives the number of legal documents $DL_i \in DL_1, DL_2, DL_3 \ldots DL_m$ collected from the dataset. After receiving the input legal documents, the keyword extraction process is carried out by applying the Sammon keyword mapping technique. It is a machine learning technique that maps the keywords from the high-dimensional space into a lower dimensionality based on the distance measure.

$$D_{ij} = \sum \sum \left(S_{Freq} - \delta\right)^2 \tag{1}$$

where, D_{ij} denotes a distance between the score of the frequent occurrence of the word 'S_{Freq}', δ indicate a threshold. The word frequency score is measured as the ratio of the number of times the words appear in the given document to the total number of words in the documents.
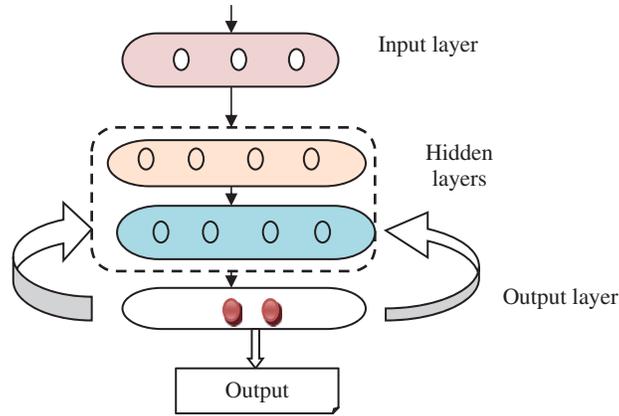

Figure 2: Structural design of recurrent multilayer perceptive deep neural classification

$$S_{Freq} = \left(\frac{v_t(DL)}{v_k}\right) \tag{2}$$

From (2), S_{Freq} indicates a frequent occurrence score of the word, $v_n(DL)$ indicates the number of times the words appeared in the document, 'v_k' denotes a total number of words in the specified document. From (1), the frequent occurrence score of the word closer to the threshold is selected as an important keyword. This process increases accuracy and minimizes time consumption.

After performing the keyword extraction, classification is performed using Quadratic discriminant analysis. The Quadratic discriminant analysis is a classifier that helps to separate the input into two or more classes based on the likelihood ratio test. The likelihood ratio test is performed between the legal documents and the mean of their respective class. The likelihood is estimated as given below,

$$L\langle DL_i|g_j\rangle = \sqrt{(2\pi V)^{-n}} \exp\left[-0.5 * \sum_{i=1}^{m}\left(\frac{(DL_i - g_j)^2}{V^2}\right)\right] \tag{3}$$

From (3), $L\langle DL_i|g_j\rangle$ specifies the likelihood between the document ($DL_i \in DL_1, DL_2, DL_3 \ldots DL_m$ and mean (g_j), 'V' specifies a standard deviation. The likelihood ratio offers the results from '0' to '1'. It also expresses that how the documents are more related to the mean of the particular class. As a result, the likelihood function classifies the documents into different classes. In other words, the document which is closer to the mean of the particular class is classified. In this way, all the documents are correctly classified and minimize incorrect classification. After classifying, the training error is calculated to minimize the false positive rate. The training error is measured as the squared difference between the actual and predicted classification results. The error rate is formulated as given below,

$$e = [y_i - y_o]^2 \tag{4}$$

where 'e' indicates a training error, y_i indicates an actual output, y_o represents the predicted output. Then the process is repeated until the proposed SKM-QDRMPDNC technique finds the minimum training error in the document classification. The classification result with minimum training error is considered as final result. Hence the proposed technique is called a recurrent neural network. This process minimizes the incorrect classification and enhances the accuracy with minimum time complexity. The algorithmic process of the SKM-QDRMPDNC technique is described as given below

---

Algorithm 1: Sammon Keyword Mapping based Quadratic Discriminant Recurrent Multilayer Perceptive Deep Neural Classification

---

Input: Dataset, number of legal documents $DL_i \in DL_1, DL_2, DL_3 \ldots DL_m$
Output: Increase classification accuracy
Begin
Collect the number of legal documents '$DL_i = DL_1, DL_2, DL_3 \ldots DL_m$' collected from dataset —-
[input layer]
   For each collected legal document 'DL_i'
      Apply Sammon mapping —[**hidden layer 1**]
         For each word in the document
            Compute frequency score S_{Freq}
            Measure the distance 'D_{ij}'
            Select the significant keywords
         end for
      Initialize the number of classes $c_1, c_2, c_3 \ldots c_k$ and mean 'g_j' —**hidden layer 2**
         For each document 'DL_i' with selected keywords 'v_i'
         for each mean g_j
            Measure the likelihood ratio test '$L\langle DL_i|g_j\rangle$'
          If the document is closer to mean 'g_j' **then**
            Classify the legal documents into a particular class
          End if
            The process is repeated until minimum error is obtained
          return (classified legal documents) at the **output layer**
          end for
      end for
End

The above algorithm provides the step-by-step process of legal document classification based on keyword extraction and classification. Initially, the number of legal documents is collected from the dataset. Then the keyword extraction is performed using Sammon mapping based on the distance measure. The frequency score relevant to the threshold is selected as a significant keyword for classifying the documents. After the feature extraction, the classification is done at the second hidden layer. Depending on distance measures machine learning technique applies machine keywords from high-dimensional space into lower dimensionality. The Quadratic Discriminant classifier is applied for measuring the likelihood between the documents and the mean of the classes. Based on the likelihood measure, the legal documents are correctly classified at the output layer with higher accuracy. Likelihood function classifies the documents into different classes. The document which is closer to the mean of a particular class is classified. The process is repeated until the proposed SKM-QDRMPDNC technique finds the minimum training error in the document classification. The classification result with minimum training error is considered as final result.

3 Experimental Settings

In this section, experimental assessment of the proposed SKM-QDRMPDNC technique and existing methods namely JBLACN [1] and hierarchical nested attention structure model [2] are implemented using Legal Case Reports Data Set taken from the UCI machine learning repository https://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports. This dataset contains Australian legal cases collected from the Federal Court of Australia (FCA). This includes all cases from the years 2006, 2007, 2008, and 2009. The dataset which consists of the court report documents have been used to perform automatic legal citation classification. For each document in the dataset, catchphrases, citations sentences, citation catchphrases, and citation classes are collected. Citation classes are indicated in the document, and denote the type of treatment given to the cases cited by the present case.

For each case, a list of labeled citations are provided. A case may be cited differently in different citing cases or even within the same citing case. This is due to the fact that different aspects of the cited case may be of interest.

The distribution among the citation classes [26] for the years 2006–2009 from 3890 FCA documents containing 18715 labelled citations is shown in Tab. 1.

Table 1: Distribution of citation classes for 2006–2009

Applied	1803	Referred to	3017	Followed	1759	Cited	9346
Distinguished	463	Discussed	706	Related	94	Considered	1339
Approved	61	Quoted	87	Not Followed	57	Affirmed	91
Explained	10	Ref to	15	Questioned	9	Reversed	20
Relied on	4	Noted	7	Doubted	3	Disapproved	8
Overruled	2	Adopted	2	Referred	2	Compared	2

The dataset consists of around 3890 documents. Among the 3890 documents, the number of documents taken for experimental analysis are in groups of 300, 600, 900…3000. The proposed technique is applied on the input set of documents and results in classifying documents based on the citation classes listed. This task of classifying documents is vital to law professionals to know whether the decision has received positive, negative, cautionary or neutral treatment in subsequent judgements.

4 Performance Analysis

The performance evaluation of the SKM-QDRMPDNC technique and existing methods namely JBLACN [1] and hierarchical nested attention structure model [2] are done with certain parameters such as accuracy, precision, recall, F-measure, and computational time. The performance results of proposed and existing techniques are discussed as given below,

4.1 Impact of Accuracy

Accuracy is measured as the ratio of numbers of legal documents that are correctly categorized into various classes to the total number of documents. The accuracy of the three methods are estimated as follows,

$$Acc = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} * 100 \tag{5}$$

From (5), Acc indicates accuracy, T_p indicates the true positive, T_n specifies the true negative, F_p symbolizes the false positive, F_n indicates a false negative. The overall accuracy is measured in terms of percentage (%).

Tab. 2 provides the experimental outcome of the accuracy of three methods namely the SKM-QDRMPDNC technique and existing methods namely JBLACN [1] and hierarchical nested attention structure model [2] *vs.* several legal documents taken in the ranges from 300 to 3000. The accuracy is measured based on the number of legal documents correctly identified. The comparative analysis indicates that the SKM-QDRMPDNC technique offers improved performance in terms of achieving higher accuracy of legal documents classification. As shown in Tab. 2, '300' numbers of legal documents are considered for conducting the experiments and the accuracy is 91% using the SKM-QDRMPDNC and the accuracy is 87.33% and 85% using the conventional methods JBLACN [1] and hierarchical nested attention structure model [2] for each run. Similarly, other results are observed with the varying number of legal documents. From the comparative analysis, the accuracy is found to be higher using the SKM-QDRMPDNC technique than JBLACN [1] and the hierarchical nested attention structure model [2]. The average of ten assessment results of the SKM-QDRMPDNC technique shows that the accuracy is found to be improved by 4% when compared to [1] and 6% when compared to [2] respectively.

4.2 Impact of Precision

It is measured as the ratio of relevant documents that are correctly classified to the total number of documents. The precision of different algorithms is expressed as given below,

$$PR = \left(\frac{T_p}{T_p + F_p} \right) * 100 \tag{6}$$

From (6), PR indicates precision, Tp indicates a true positive, Fp indicates a false positive. The precision is measured in terms of percentage (%).

Table 2: Accuracy

Number of legal documents	Accuracy (%)		
	SKM-QDRMPDNC	JBLACN	Hierarchical nested attention structure model
300	91	87.33	85
600	92.5	89.66	87.16
900	94.44	90	88.33
1200	94.16	89.16	85.83
1500	93.86	90.66	88
1800	93.61	90	87.77
2100	92.38	89.5	87
2400	91.66	88.75	86.66
2700	90.37	88.14	86.29
3000	90.33	88.33	86.04

Fig. 3 demonstrates the precision for different numbers of legal documents in the range of 300 to 3000. From the graphical design, the precision is notably increased using the SKM-QDRMPDNC technique than the conventional methods. The precision of the SKM-QDRMPDNC technique is 94.29%. The precision of JBLACN [1] and hierarchical nested attention structure model [2] are 92.06% and 89.79%. This is due to the application of the deep learning technique for accurately analyzing the extracted keywords with the legal documents collected from the dataset. The likelihood ratio test indicates that the SKM-QDRMPDNC technique accurately classifies the documents and minimizes the incorrect classification. From the analysis, the precision is found to be increased using the SKM-QDRMPDNC than JBLACN [1] and hierarchical nested attention structure model [2]. The average comparison results indicate that the SKM-QDRMPDNC technique indicates that the precision is found to be improved by 2% when compared to [1] and 4% when compared to [2] respectively.
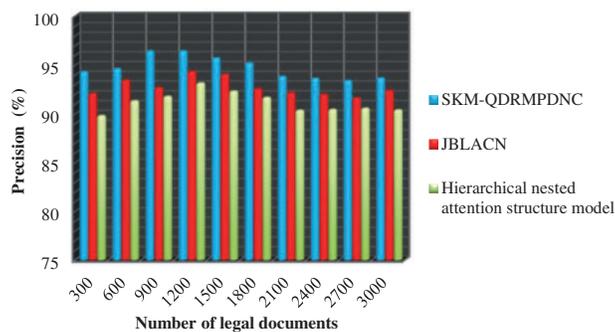
Figure 3: Performance analysis of precision

4.3 Impact of Recall

It is defined as the ratio of relevant documents that are correctly classified to the total number of relevant documents. Therefore, the overall recall rate is measured as given below,

$$RR = \left(\frac{Tp}{Tp + Fn} \right) * 100 \tag{7}$$

From (7) 'RR' denotes a recall, Tp denotes true positive, Fn specifies a false negative. The recall rate is measured in terms of percentage (%).

Table 3: Recall

Number of legal documents	Recall (%)		
	SKM-QDRMPDNC	JBLACN	Hierarchical nested attention structure model
300	95.38	92.8	91.66
600	97.23	94.85	93.66
900	97.59	96.20	94.77
1200	97.29	93.45	90.47
1500	97.66	95.45	93.75
1800	97.83	96.12	94.07
2100	97.84	95.91	94.64
2400	97.15	95.12	94.02
2700	95.76	94.78	93.36
3000	95.43	94.18	93.25

Tab. 3 explains the performance results of recall rate *vs.* several legal documents collected from the dataset. The numbers of legal documents are taken as an input in the ranges 300 to 3000 for calculating the recall. The recall rate is measured based on the true positive and false negatives. As shown in Tab. 3, the SKM-QDRMPDNC technique minimizes the recall rate of document classification when compared to existing techniques. For instance, with 300 legal documents, the true positive and false negative of the SKM-QDRMPDNC technique are 248 and 12. Whereas the true positive and false negative of existing JBLACN [1] are 232 and 18 and true positive and false negative of existing hierarchical nested attention structure model [2] are 220 and 20. Therefore, the overall recall rate was found to be '95.38%', '92.8%' and '91.66%' using the SKM-QDRMPDNC technique, [1], and [2] respectively. The above arithmetic analysis confirms that the recall rate is considerably increased using the SKM-QDRMPDNC technique. The performance results of SKM-QDRMPDNC are compared to existing methods. The average of comparison results reveals that the proposed technique increases the recall rate by 2% when compared to [1] and 4% when compared to [2].

4.4 Impact of F-measure

It is measured based on the mean of precision as well as recall. It is formulated as given below,

$$F - measure = 2 * \left(\frac{PR * PR}{PR + PR} \right) * 100 \tag{8}$$

From (8), PR denotes precision, 'RR' denotes a recall. F-measure is measured in terms of percentage (%).

Fig. 4 illustrates the performance results of the F-measure to the number of legal documents. The F-measure is estimated based on precision and recall. The tabulated results reveal that the proposed SKM-QDRMPDNC technique outperforms well in terms of achieving higher F-measure. Let us consider '300' numbers of legal documents for conducting the experiments and the number of F-measure is 94.83% using the SKM-QDRMPDNC and F-measure is 92.42% and 90.71% using the conventional methods JBLACN [1] and hierarchical nested attention structure model [2]. The overall performance of the F-measure using the proposed SKM-QDRMPDNC technique is compared to other existing methods. The overall comparative results indicate that the F- Measure of the proposed SKM-QDRMPDNC technique significantly increased by 2% and 4% than the existing JBLACN [1] and hierarchical nested attention structure model [2]. This is because the SKM-QDRMPDNC technique achieves higher precision and recall rate.
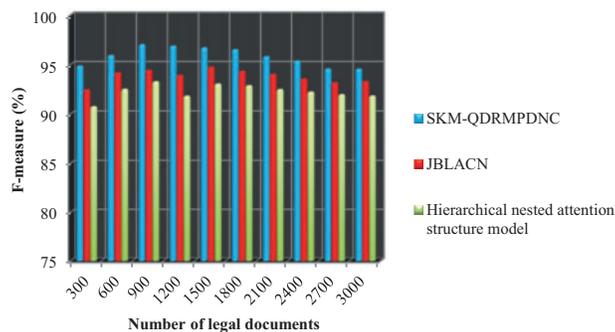

Figure 4: Performance analysis of F-measure

4.5 Impact of Computational Time

It is formulated as the amount of time consumed by the algorithm for categorizing the documents into different classes. As a result, the overall consumption of the time is measured as given below,

$$CT = n * Time[CSD] \tag{9}$$

From (9), CT shows a computational time, n denotes the number of legal documents, CSD specifies a classification of a single document. The overall computational time is measured in terms of milliseconds (ms).

Tab. 4 indicates the performance results of computational time for finding the document's classification time. From the observed results, the SKM-QDRMPDNC technique decreases the computational time when compared to existing classification techniques. For example, with 300 legal documents, the computational time of a single document is '0.074 ms' using the SKM-QDRMPDNC technique, whereas the computational time is '0.08 ms' using hybrid [1] and time for classifying the one document being '0.09 ms' using [2]. Therefore, the overall computational time was found to be

'22 ms', '24 ms', and '27 ms' using the SKM-QDRMPDNC technique, [1], and [2] respectively. The above numerical analysis validates that the computational time is considerably decreased using the SKM-QDRMPDNC technique than the other two existing methods.

Table 4: Computational time

Number of legal documents	Computational time (ms)		
	SKM-QDRMPDNC	JBLACN	Hierarchical nested attention structure model
300	22	24	27
600	27	30	33
900	30	32	36
1200	34	36	42
1500	36	41	45
1800	40	43	47
2100	44	48	50
2400	48	50	53
2700	51	54	57
3000	54	57	60

The average of ten comparative results exposes that the SKM-QDRMPDNC technique decreased the document classification time by 7% when compared to [1] and 15% when compared to [2].

5 Conclusion

In this paper, a novel deep learning technique called SKM-QDRMPDNC is proposed for learning the textual legal document representations for automatic legal citation classification. This task characterizes the relation between the present case and the cited ones, which is a novel application in itself helping law professionals in better understanding of court decisions. The proposed technique receives input legal documents and keyword extraction process is performed by Sammon keyword mapping technique. Depending on distance measure, machine learning technique maps the keywords from high-dimensional space into lower dimensionality. After keyword extraction, Quadratic Discriminant analysis is applied into a hidden layer of recurrent multilayer perceptive deep neural network to perform the legal document classification based on likelihood measure. Lastly, classification results are obtained with higher accuracy. The classification result with minimum frequency error is considered as final class result. Experimental result is performed with various parameters namely accuracy, precision, recall, F-measure, and computational time *vs.* several legal documents collected from the dataset. The proposed technique performs legal document classification, however time complexity involved in the legal document data analytics could be further reduced. In future work, SKM-QDRMPDNC technique is extended for learning the textual legal document representations by using optimization based deep learning methods.

References

[1]   D. Ji, P. Tao, H. Fei and Y. Ren, "An end-to-end joint model for evidence information extraction from court record document," *Information Processing & Management, Elsevier*, vol. 57, no. 6, pp. 1–14, 2020.

[2]   K. Zhu, B. Ma, T. Huang, Z. Li, H. Ma *et al.,* "Sequence generation network based on hierarchical attention for multi-charge prediction," *IEEE Access*, vol. 8, pp. 109315–109324, 2020.

[3]   M. Sharma and M. Bilgic, "Learning with rationales for document classification," *Machine Learning, Springer*, vol. 107, no. 5, pp. 797–824, 2018.

[4]   S. Sharafat, Z. Nasar and S. W. Jaffry, "Data mining for smart legal systems," *Computers & Electrical Engineering, Elsevier*, vol. 78, no. December(6), pp. 328–342, 2019.

[5]   Y. Fang, X. Tian, H. Wu, S. Gu, Z. Wang *et al.,* "Few-shot learning for chinese legal controversial issues classification," *IEEE Access*, vol. 8, pp. 75022–75034, 2020.

[6]   W. Zhang, Y. Li and S. Wang, "Learning document representation via topic-enhanced LSTM model," *Knowledge-Based Systems, Elsevier*, vol. 174, pp. 194–204, 2019.

[7]   C. Du and L. Huang, "Text classification research with attention-based recurrent neural networks," *International Journal of Computers Communications & Control*, vol. 13, no. 1, pp. 50–61, 2018.

[8]   G. Li, Z. Wang and Y. Ma, "Combining domain knowledge extraction with graph long short-term memory for learning classification of chinese legal documents," *IEEE Access*, vol. 7, pp. 139616–139627, 2019.

[9]   C. I. Hausladena, M. H. Schubert and E. Ash, "Text classification of ideological direction in judicial opinions," *International Review of Law and Economics, Elsevier*, vol. 62, no. 10, pp. 1–19, 2020.

[10] Y. Ma, P. Zhang and J. Ma, "An ontology driven knowledge block summarization approach for chinese judgment document classification," *IEEE Access*, vol. 6, pp. 71327–71338, 2018.

[11] L. Liu, K. Liu, Z. Cong, J. Zhao, Y. Ji *et al.,* "Long length document classification by local convolutional feature aggregation," *Algorithms*, vol. 11, no. 8, pp. 1–12, 2018.

[12] I. S. Bajwa, F. Karim, M. A. Naeem and R. U. Amin, "A semi supervised approach for catchphrase classification in legal text documents," *Journal of Computers*, vol. 12, pp. 451–461, 2017.

[13] S. Li, H. Zhang, L. Ye, X. Guo and B. Fang, "MANN: A multichannel attentive neural network for legal judgment prediction," *IEEE Access*, vol. 7, pp. 151144–151155, 2019.

[14] C. V. Trappey, A. J. C. Trappey and B. H. Liu, "Identify trademark legal case precedents - Using machine learning to enable semantic analysis of judgments," *World Patent Information, Elsevier*, vol. 62, no. 1, pp. 1–10, 2020.

[15] R. A. Shaikha, T. P. Sahua and V. Anand, "Predicting outcomes of legal cases based on legal factors using classifiers," *Procedia Computer Science, Elsevier*, vol. 167, pp. 2393–2402, 2020.

[16] T. M. Paixao, R. F. Berriel, M. C. S. Boeres, A. L. Koerich, C. Badue *et al.,* "Self-supervised deep reconstruction of mixed strip-shredded text documents," *Pattern Recognition, Elsevier*, vol. 107, no. 7, pp. 1–37, 2020.

[17] Y. Zhang, M. Tuo, Q. Yin, L. Qi, X. Wang *et al.,* "Keywords extraction with deep neural network model," *Neurocomputing Elsevier*, vol. 383, no. 2, pp. 113–121, 2020.

[18] M. N. Awan and M. O. Beg, "TOP-Rank: A topical postion rank for extraction and classification of keyphrases in text," *Computer Speech & Language, Elsevier*, vol. 65, pp. 1–29, 2021.

[19] Z. Huang and Z. Xie, "A patent keywords extraction method using Text Rank model with prior public knowledge," *Complex & Intelligent Systems, Springer*, vol. 7, pp. 1–12, 2021.

[20] S. Susan and J. Keshari, "Finding significant keywords for document databases by two-phase maximum entropy partitioning," *Pattern Recognition Letters, Elsevier*, vol. 125, pp. 195–205, 2019.

[21] D. Song, A. Vold, K. Madan and F. Schilder, "Multi-label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training," *Information Systems*, vol. 106, pp. 101718, 2021.

[22] D. Ji, J. Gao, H. Fei, C. Teng, Y. Ren *et al.,* "A deep neural network model for speakers coreference resolution in legal texts," *Information Processing & Management*, vol. 57, no. 6, pp. 102365, 2020.

[23] M. J. Tsai, Y. H. Tao and I. Yuadi, "Deep learning for printed document source identification," *Signal Processing: Image Communication*, vol. 70, pp. 184–198, 2019.

[24] L. R. C. de Mendonça and G. da. C. Júnior , "Deep neural annealing model for the semantic representation of documents," *Engineering Applications of Artificial Intelligence*, vol. 96, no. 6, pp. 103982, 2020.

[25] A. Quemy and R. Wrembel, "ECHR-OD: On building an integrated open repository of legal documents for machine learning applications," *Information Systems*, vol. 106, pp. 1–20, 2021.

[26] F. Galgani and A. Hoffmann, "Lexa: Towards automatic legal citation classification," in *AI 2010: Advances in Artificial Intelligence*, J. Li (eds.), Vol. 6464, pp. 445–454, 2010.