A Study on Legal Judgment Prediction using Deep Learning Techniques

Prameela Madambakam
Department of Computer Science and Engineering
Indian Institute of Information Technology, Sri City
India
prameela.8cse@gmail.com

Shathanaa Rajmohan
Department of Computer Science and Engineering
Indian Institute of Information Technology, Sri City
India
shathanaaraj@gmail.com

2022 IEEE Silchar Subsection Conference (SILCON) | 978-1-6654-7100-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SILCON55242.2022.10028879

***Abstract*— Legal Judgment Prediction (LJP) involves examining the given input case document and recommending the judgment prediction such as applicable law sections, charges, and penalties as delivered by the judge in the court. It assists the judges and lawyers in analyzing and resolving the given case. The various steps involved in LJP equip the lawyers with supporting points to argue the case in the court and the parties involved with the probability of winning the case by predicting the judgment outcome. This paper surveys recent state-of-the-art LJP algorithms published between 2018 and 2022 by focusing on various factors such as Deep Learning (DL) and Artificial Intelligence (AI) ambient techniques, civil and criminal case types, evaluation measures, various data sets available, prediction and modelling methods, challenges, and limitations. Based on this study we derived a taxonomy that will organize the collected papers into two channels called criminal and civil cases which are further classified based on the techniques used for prediction.**

Keywords— Legal judgment prediction, Multi-task prediction, Civil cases, Criminal cases, Deep learning, Text classification, Legal reading comprehension

I. INTRODUCTION

Legal or judicial judgment prediction is an assistance system to recommend the judicial decision components called charges, terms of penalty, and the law sections depending on the case facts. As per statistics [1], more than 3 crore pending cases exists in various courts in India for more than 10 years. The approximate count of these cases is 60,000 in the supreme court, 42 lakhs in high courts, and 2.7 crores in the district and subordinate courts. Out of 5 crore new cases being registered every year, judges can dispose only 2 crore cases. The rationale behind is: 1. increasing awareness of the rights of the common man 2. new mechanisms and rights such as Public Interest Litigation (PIL) and Right to Information (RTI) 3. lack of sufficient judges and courts 4. too much litigation from the government 5. poor quality results in lower courts 6. ambiguous draft of laws. It leads to the development of LJP technology to assist the judges in the case study, deliver judgment on time, and speed up the judicial system digitally.

The objective of LJP algorithms is to perform two functions [2]: (1) predicting judgment to assist adjudicators and (2) avoiding the delivery of a wrong verdict. It acts as an alert system for discriminations between predicted and delivered judgments in real-time.

Based on a survey [3], it was found that only 20% of judges exist for every ten lakh citizens in India. Application of these legal frameworks in India such as the Alternative Dispute Resolution (ADR) mechanism / out-of-the-court settlement by lawyers or law firms will greatly reduce the cases and accelerate the delivery of justice [4][5]. Following is the summary of the benefits of using DL in LJP: 1. expedite the prediction task and results 2. spontaneous validation of the case data 3. quick access to the precedents 4. easy to identify the corruption by noticing the great variations in human and predicted judgment.

A lot of works has been done to solve LJP using Natural Language Processing (NLP), Reinforcement Learning (RL), DL and attention mechanisms, first-order logic, cognitive computing, and soft computing techniques. But there is no survey paper published on recent works concentrating on DL and attention mechanisms to solve criminal and civil cases which are discussed in this work.

The body of this article is structured as follows: section 2 explains the taxonomy, section 3 discusses the literature review, section 4 lists the legal datasets, and section 5 discusses the conclusion & future scope.

II. TAXONOMY OF DL-BASED LEGAL JUDGMENT PREDICTION

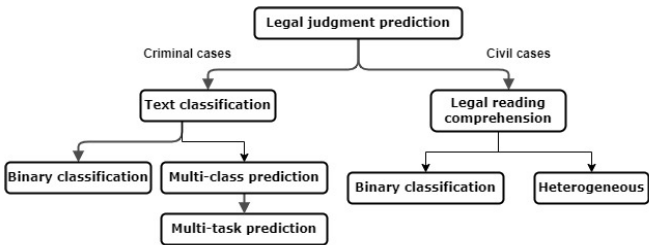

Fig. 1: The taxonomy of legal judgment prediction

This work studies the top-tier publications between 2018 and 2022 that solve the LJP and categorize the collected papers in taxonomy as criminal and civil cases as shown in Fig. 1.

We found that judgment in the *criminal cases* were predicted using text classification techniques which are

978-1-6654-7100-8/22/$31.00 ©2022 IEEE

categorized into binary and multi-class predictions based on the possible potential classes in the predicted outcome. Multi-class prediction can be further classified as a multi-task prediction based on the number of tasks identified and predicted in the outcome of a case. In general, criminal cases involve multiple tasks and the resultant final judgment is classified as one among the precedent results for law article and charge prediction tasks. Alternatively, the penalty prediction can be made based on regression techniques.

The text classification technique applied to criminal cases involves two modules: case modeling and prediction modules.

Case modeling module: It takes up the task of converting the complex unstructured legal case documents into structured documents based on which judgment prediction is made in the subsequent prediction module using Neural Networks (NN) or regression techniques [6]. Some existing case models are [7]: unsupervised tensor decomposition, tensor models, feature models, and matrix decomposition. Other forms of encoding include word vectors [8], and first-order logic [9].

Prediction module: Available prediction algorithms fall into regression algorithms and sequential neural networks. Existing works use different variations of NN including [7], Text Convolutional Neural Networks (TextCNN), TextCNN attention, Text Recurrent Neural Network (TextRNN), TextRNN attention, Gated Recurrent Unit (GRU), Bidirectional GRU (Bi-GRU), Long Short Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM) are used to extract the important components from a legal document that helps in decision prediction [10]. Regression algorithms used are classified into two: 1) basic Machine Learning (ML) techniques and 2) customized regression techniques. Basic ML-based techniques include [7]: ElasticNet regression, linear regression, support vector regression, and polynomial regression. Customized techniques include 1. optimized oloss regression technique [7] - combination of tensor decomposition as case modeling and lasso regression technique. 2. optimized oridge regression technique [11] - a combination of tensor decomposition modeling and ridge regression prediction technique.

The benefit with text classification techniques is their lower time complexity during training. But it suffers from the drawback of ignoring infrequent law articles which occurs as a result of neglecting text semantics and depending highly on statistics. To overcome this drawback some text classification techniques are designed as a sequence prediction problem using deep learning techniques [8] which can be one of the following:

Convolutional Neural Networks (CNN): It acts as a classifier to learn the important case components and merge the result with law articles without counting the relation between the law articles into account. Some articles have dependencies between them i.e., discusses about the same crime and both can be cited parallelly, whereas the other articles conflict with each other and avoid occurring in the same document.

Sequence-to-sequence based: It overcomes the drawback of CNN by considering the relations between the law articles in addressing the crime given in the fact description and acts as a sequence predictor using an attention mechanism.

The *civil cases* are solved mainly using Legal Reading Comprehension (LRC) technique which extracts important components from the legal text and is taken as an advantage to predict the legal judgment. Judgment prediction in civil cases depends on fact description, pleas of the plaintiff, and related law articles, whereas criminal cases are based on the respective precedents. Therefore, judgment predictions in civil cases may not be binary always but can be manifold and flexible such as compensation amount. Accordingly, civil cases are further categorized into binary and heterogeneous based on the outcome predicted. In binary, output predicted is one of the two possible cases like in a divorce, case proceedings, and bail cases. Alternatively, in the heterogeneous category predicted outcome vary from case to case based on the input case document, pleas, and law sections. More generally, LRC is like a question-answering system that will take plaintiff pleas (p), fact description (f), and law articles (l) then deliver answer (r) from law sections (s) and formalized as <f, p, l, r> [6].

III. LITERATURE REVIEW

We organized the collected papers as per the taxonomy defined in Fig.1 as follows. The summary of these collected papers is listed in Table 1.

3.1 Criminal Cases- Text Classification

3.1.1 Binary classification

Pillai et al. [12] adapted the bag-of-words technique, an NLP tool to analyze the court proceedings documents by extracting key terms. It later applied CNN to classify the case into one of the predefined charges to predict an offense as either non-bailable or bailable. Malik et al. [13] created an Indian Legal Document Corpus (ILDC) dataset with Indian supreme court case proceeding documents for case acceptance decision. It was evaluated using various DL models for judgment prediction and found that XLNet + Bi-GRU hierarchical model provided the highest performance.

3.1.2 Multi-class prediction

Theresa et al. [14] proposed a composite technique by combining a dynamic fuzzy neural network and Maximum Spanning Tree (MST) for the classification of criminal cases called murder cases. Case documents are modeled, trained, and tested one at a time as n feature vector represented as MST and predict the three-class output called charge, penalty, and applicable laws. Li et al. [15] introduced a Multiple channel Attention Neural Network Model (MANN) to execute LJP jobs using precedents in a unified framework. This work adopted two-tier attention-based sequence encoders to design the lingual conversation from distinct sections of the input document at the sentence as well as word levels by utilizing the Bi-GRU model.

Thomas et al. [16] derived semi-supervised pattern-based mining, a knowledge-driven method to retrieve domain-knowledge features by mining the legal data. This was accomplished by tagging legal terms in the corpus using the Judicial Case Ontology (JCO). Chen et al. [17] applied deep learning models to predict the judgment outcome by building three models named penalty forecasting using TextCNN and FastText, a judicial provision forecasting using TextCNN, and an allegation forecasting using FastText and TextCNN.

Jin et al. [18] proposed a methodology using term fuzzification, document vector regression, and information extraction on drug-related judgment reports to forecast the intended prison sentence using fuzzy text and value-based rules. Zhu et al. [2] addressed a novel hierarchical nested attention structure model by highlighting the correlation between distinct charges. They adapted the encoder at sentence-level and word-level attention, and the decoder with the LSTM attention approach for charge prediction.

Guo et al. [11] derived the TenRR model comprising tensor decomposition as case modeling and ridge regression as a prediction technique for verdict prediction of legal cases. It is composed of three phases named RTenr, Itend, and ORidge. Xiao et al. [19] addressed a pre-trained language model called lawformer, as well as an encoder using the full self-attention technique. It blends global attention and dilated sliding window attention to encode lengthy sequences by capturing the long-distance dependencies. It was assessed on some standard law activities, including precedents retrieval, judgment forecasting, legal question answering, and LRC.

Guo et al. [7] proposed the TenLa model by combining the case modeling method of controllable tensor decomposition and a prediction method of optimized lasso regression model using commonality between the documents. This architecture is composed of 3 modules named MonTen, ConTen, and OLass. Xu et al. [20] developed a Law Article Distillation-based Attention Network (LADAN), and Graph Distillation Operator (GDO) to extract the differences between ambiguous law sections and charges. It also developed an innovative attention approach to learn compelling distinguishing features from fact descriptions attentively.

Chen et al. [8] addressed data imbalance using weight sharing classification and missing value concerns using nonstatic word embedding methods in law text analysis for law article prediction. Word vectors act as case modeling and KNN classifier, TF-IDF algorithm, and sequence-to-sequence algorithms are applied as prediction algorithms. Li et al. [9] modeled the judicial theory to improve the interpretability by adapting a dual-layer criminal mechanism to model sentence forecasting technique. It is a conventional crime constitution conjecture formed with two layers called objective illegality and subjective responsibility depending on the case facts used by the judge during the decision.

Li et al. [21] addressed a law Article Deduplication Attention Network (ADAN), a new aggregation approach to distinguish confusing charges and improved semantic representations of fact description using Hierarchical Attention Mechanism (HAN) and Bi-GRU. Yue et al. [22] established a circumstance-knowing judicial judgment prediction method named NeurJudge. It was later enhanced to NeurJudge+ by employing the graph label embedding method to embed the rhetorical labels such as law sections and charges and then produce the descriptive fact representations of ambiguous verdicts.

3.1.2.1 Multi-task prediction

Zhong et al. [4] derived a topological multi-task extractor TOP JUDGE, by incorporating Directed Acyclic Graph (DAG) dependencies among multiple subtasks for judgment prediction. Huang et al. [23] developed a streamlined text-to-text transformer where sub-task dependencies were generated by the auto-regressive decoder for LJP. The two new sub-tasks named law article prediction and court view generation offer intelligible explanations for model outputs.

Yao et al. [24] explored three properties called Commonalities, Specificities, and Dependencies abbreviated as CSDNet, and their relationships among the various subtasks involved in LJP. It was done using learning, denoising, and reinforcement modules. Yao et al. [25] derived a new GHE - DAP judicial forecasting model to find various subtasks simultaneously called charge, penalty term, and law article prediction. Here, Gated Hierarchical Encoder (GHE) extracts absolute rhetorical information of case facts and dependencies. Auto-learning predictor (DAP) dynamically learns the reliance between subtasks. Lyu et al. [26] introduced the Criminal Element Extraction Network (CEEN), a RL model to address the complications raised by ambiguous law articles. This work employed a RL - based extractor and a multi-task detector for judgment forecasting.

3.2 Civil Cases –Legal Reading Comprehension

3.2.1 Binary classification

Shangbang et al. [6] derived a new LRC model Auto-Judge, to extract the in-depth rhetorical conversations among laws, facts, and pleas using a text encoder, pair-wise attentive reader, and an output module. Branting et al. [5] explain two strategies for predicting explainable judicial decisions listed: 1. attention weights and attention network to focus on salient case text, and 2. Semi-supervised Case Annotation for Legal Explanations (SCALE) with both predictable relationships and explanatory value.

Karl et al. [27] described two mechanisms for reasonable legal decision listed: Attention Network-based Prediction (ANP) for decision support and SCALE by exploiting the semantic and structural regularities in the text. Li et al. [28] derived a cognitive computing system for anticipating legal judgments through three-layer architecture termed as knowledge learning layer, semantics understanding layer, and knowledge reasoning layer. This work applied CNN and Bi-LSTM to perform legal information augmentation.

Malik et al. [13] created ILDC with the last 512 tokens of Indian supreme court case proceedings labeled with original decisions with three types of datasets: $ILDC_{multi}$, $ILDC_{Single}$, and $ILDC_{expert}$. Evaluated ILDC on various DL models such as classical, sequential, transformer, hierarchical with attention for judgment prediction, and hierarchical occlusion model for the explanation.

3.2.2 Heterogeneous

Zhou et al. [29] developed an innovative multi-task learning framework to resolve the disputes of e-commerce transactions through Legal Dispute Judgment (LDJ) prediction. It focused on the evidence of the ongoing transaction, historical behavior data of the buyer, and seller. Zhong et al. [30] addressed a RL method named QAjudge, based on the two principles: avoid predicting too many elements and then apply the RL reward function to ask minimum questions to extract important components supporting prediction.

TABLE 1: Summary of the collected papers.

S. No	Authors	Methodology (Performance)	Dataset
3.1 Text Classification – Criminal Cases			
3.1.1 Binary classification			
1.	Pillai et al. 2020 [12]	Bag-of-words+ CNN (Accuracy-85%)	Indian criminal and civil cases
2.	Malik et al. 2021 [13]	ILDC for CJPE: applied various classical, sequential, transformer, hierarchical, and attention models; Explainability: hierarchical occlusion-based model (XLNET+BiGRU: Accuracy-77.78%)	Indian supreme court case proceedings dataset named ILDC
3.1.2 Multi - class prediction			
3.	Theresa et al. 2015 [14]	Hybrid algorithm: dynamic fuzzy supervised neural network architecture + maximum spanning tree (Accuracy-97.98%)	Supreme court of India judgments, Indian penal code
4.	Li et al. 2019 [15]	MANN: Attentive sequence encoders: Bi - GRU +HAN (Accuracy-95.5%)	Refined CAIL 2018 datasets: RCAIL
5.	Thomas et al. 2019 [16]	Semi-supervised pattern-based learning + judicial case ontology (F1 score – 98.9%)	480 criminal e-judgments.
6.	Chen et al. 2019 [17]	FastText +TextCNN (F1 score – 86.27%)	CAIL 2018: CAIL - small + CAIL - big
7.	Jin et al. 2020 [18]	Value-based rules and fuzzy text + document vector regression (Root Mean Square Error: -2.008)	Drug-related criminal cases
8.	Zhu et al. 2020 [2]	Encoder using word-level attention, decoder: LSTM + attention mechanism (F1 score – 93%)	CJO, CAIL data sets
9.	Guo et al. 2020 [11]	TenRR: Unsupervised tensor decomposition + optimized ridge regression model (Accuracy-93.71%)	Chinese reference document network.
10.	Xiao et al. 2020 [19]	Lawformer pre-trained language model: global attention mechanism + dilated sliding window attention approach (MacroF1 score for classification task:72.1%)	CAIL - Long
11.	Guo et al. 2020 [7]	TenLa: Unsupervised tensor decomposition + optimized lasso regression model (Accuracy-93%)	Chinese reference document network.
12.	Xu et al. 2020 [20]	LADAN: Graph Distillation Operator (GDO)+ attention method (Accuracy-96.57%)	CAIL 2018: CAIL - small +CAIL - big
13.	Chen et al. 2021 [8]	Nonstatic word embedding methods+ weight sharing classification+ transfer learning model (Accuracy-91%)	Charge, law article, few-shot charge datasets
14.	Li et al. 2021 [9]	First order logic + double-layer criminal system (MacroF1-94.5%)	CAIL 2018
15.	Li et al. 2021 [21]	ADAN: Two-tier sequence encoder- Bi - GRU + hierarchical attention mechanism (Accuracy-93.66%)	Modified CAIL 2018 datasets: RCAIL -S and RCAIL - L.
16.	Yue et al. 2021 [22]	NeurJudge, NeurJudge+ graph label embedding method (Accuracy-92.64%)	CAIL 2018: CAIL - small + CAIL - big
3.1.2.1 Multi – task prediction			
17.	Zhong et al. 2018 [4]	TOP JUDGE+ DAG dependencies (Accuracy-86.3%)	CJO, PKU, and CAIL of China
18.	Huang et al. 2020 [23]	Unified text-to-text transformer, auto-regressive decoder (MacroF1- 48.8%)	Chinese criminal law
19.	Yao et al. 2020 [24]	LSTMs+ named QLSTM (Law article prediction accuracy-90.2%)	CAIL2018 - small and large datasets
20.	Yao et al. 2021 [25]	GHE-DAP method (Law article prediction accuracy- 89.5%)	CAIL2018 - small and large datasets
21.	Lyu et al. 2022 [26]	CEEN: sentence and word level Bi-LSTM, word-level attention, multi-task predictor (Law article prediction accuracy-97.43%)	CAIL2018 - small and large datasets
3.2 Legal Reading Comprehension – Civil Cases			
3.2.1 Binary classification			
22.	bang et al. 2018 [6]	Auto-Judge LRC tool (Accuracy-82.2%)	CJO
23.	Branting et al. 2019 [5]	Attention network + attention weights, SCALE (F1 score –94.3%)	WIPO domain name dispute cases
24.	Karl et al. 2020 [27]	Attention network+ salient case text based on attention weights, SCALE (Accuracy-79.5%)	WIPO domain name dispute cases.
25.	Li et al. 2020 [28]	Cognitive computing + Bi - LSTM and CNN + legal knowledge augmentation (F1 score –72.65%)	Divorce cases from CJO
3.2.2 Heterogeneous			
26.	Zhou et al. 2019 [29]	Legal knowledge graph, encoder, and decoder mechanisms. (MacroF1-72.6%)	Taobao's dispute and CJO - lawsuit case datasets
27.	Zhong et al. 2020 [30]	QAjudge model (Accuracy-93.3%)	CJO, PKU, and CAIL of China
28.	Ma et al. 2021 [31]	MSJudge: LSTM, Word embeddings, attention (MacroF1-73.6%)	Cases of private lending category.
29.	Song et al. 2021 [32]	LAMT-MLC: label-attention + domain-specific pre-training mechanism (Weighted macroF1 score on POSTURE50K: 80.2%)	POSTURE50K, EUROLEX57K
30.	Hwang et al. 2022 [33]	Encoder-decoder pre-trained language model (Accuracy-68.5%)	lbox open
3.3 Others			
31.	Bansal et al. 2019 [10]	Study of various DL applications for the judicial domain	-
32.	Zhong et al. 2020 [34]	Analyzed the performance of several LJP models.	C - LJP
33.	Sukanya et al. 2021 [3]	A systematic review of LJP and challenges, HAN model	-

Ma et al. [31] addressed a novel dataset simulating the real courtrooms based on the court debate data and the plaintiff's claims. This work later discriminated the claims through multi-task extraction for judgment prediction. Song et al. [32] created a novel extreme POSTURE50K dataset to fulfill the challenges of multilabel classification by adopting zipfian distribution. This work also proposed a LAMT-MLC DL model using a label-attention and domain-specific pre-training mechanism. Hwang et al. [33] created the large-scale legal Korean benchmark dataset *lbox open* containing six datasets including LJP-civil and LJP-criminal. This work also proposed *lcube,* a Korean legal encoder-decoder language model pre-trained with *lbox open* dataset for LJP.

3.3 Others

The following survey works are used for LJP problem understanding and taxonomy derivation.

Bansal et al. [10] studied the applications of DL in the legal field by mainly focusing on 3 issues listed: available data sets or sources, various legal activities explored, and not explored using DL techniques. This work concluded that DL methods such as CNNs, RNNs, LSTM, GRU, and multi-task models will consume less time, effort, and cost to solve legal tasks. Zhong et al. [34] described the applications of legal AI listed: legal question answering, similar case matching, and legal judgment prediction. Based on testing the multiple LJP models titled Fact Law, Top Judge, and gated network concluded that most models accurately forecast high-frequency charges or articles than low-frequency labels due to the differences between macro-F1 and micro-F1. Sukanya et al. [3] addressed the challenges of LJP faced by DL models with long-length case facts through empirical literature on LJP techniques, theoretical literature on text classification approaches, and the transformer-based model.

IV. DATASETS

Some standard data sets used for LJP are as follows:

4.1 Criminal Datasets for Text Classification

CAIL 2018: The Chinese AI and Law Challenge (CAIL) dataset contains criminal cases made available by the supreme court of China collected from the Chinese referee network [25]. Based on the size it has 2 versions called Large and Small as shown in the statistics in Table 2. CAIL 2018 contains only criminal cases thereby omitting the civil cases and has a short length than real-world cases, whereas CAIL - Long contains both types of cases with lengths equal to real-world cases [19].

TABLE 2: Statistics of the CAIL 2018 dataset [25]

Datasets	CAIL - Small	CAIL - Large
Penalty term	11	11
Sentence	104	180
Sections	135	229
Documents	371,721	2,494,029

CAIL - Long: It was framed with 1,099,605 civil cases and 1,129,053 criminal cases [19] with lengths equal to real-world cases. Criminal cases were labeled with charges, relevant laws, and penalty terms whereas civil cases were labeled with the relevant laws and causes of action. Details of the CAIL - Long dataset are given in Table 3.

TABLE 3: Statistics of the CAIL - Long dataset [19]

Case Type	#Law	Prison	#Case	#C	Len.
Criminal	244	0-180	115,849	201	916.57
Civil	330	-	113,656	257	1286.88

In the above table 3 # Case, #C, # Law, Len., Prison indicates the number of cases, count of charges or cause of actions, count of relevant laws, the average length of the input case facts, range of term penalty in months, respectively.

CJO: This dataset contains criminal cases collected from Chinese Judgements Online (CJO) [4].

PKU: It is built by extracting criminal cases from Peking University Law Online (PKU) [4].

Other data sets used are the Supreme Court Database (SCDB) provided by the university school of law, german legal corpora, lawsents, vietnamese legal dataset, and Japan legal code [10].

4.2 Civil Datasets for Legal Reading Comprehension

CJRC: Chinese Judicial Reading Comprehension (CJRC) is a renowned dataset [19] for LRC printed in the Chinese AI and law challenge contest in 2020. It contains question-answering text pairs along with evidence where each question can have 3 possible answers listed: yes/no, unanswerable, and span of words. Table 4 explores the CJRC dataset.

TABLE 4: Statistics of the CJRC dataset [19].

#Doc	#Que.	Len.	#S-Que	#U-Que	YN-Que
9532	9532	441.04	6692	948	1892

In the above Table 4 # Doc, # Que., Len., # YN-Que., # S-Que., and # U-Que. specifies the count of case documents, total count of questions, average length of the documents, questions with yes/no answer type, questions with answer type as span of words, and questions with unanswerable type, respectively.

4.3 Indian Legal Databases for Legal Research

A democratic country like India does not have the standard data sets available for various tasks of legal research. Indian authors have been using the gold standard datasets annotated manually by legal practitioners which are made available by the various legal information databases. Some reputed legal information databases named manupatra [35], indiankanoon [36], etc. provides the most exhaustive collection of judgments & orders, circulars, notifications, statutes (acts), legislative, regulatory, and procedural content covering India and international jurisdictions using AI, ML, and NLP for information retrieval.

V. CONCLUSION & FUTURE SCOPE

In this work, we derived a taxonomy with judicial DL techniques proposed to solve the criminal and civil cases published between 2018 and 2022. These mainly fall into text classification for criminal cases and LRC for civil cases and corresponding sub-classifications respectively. Based on the survey we conclude that most of the works contributed to criminal cases and very few works focused on civil cases because of the complexity of heterogeneous input documents. Decision in civil cases varies from case to case based on the case description, plaintiff pleas, and related law articles whereas criminal cases were solved using respective precedents. Therefore, we defined a category named heterogeneous containing DL models for non-binary classification irrespective of heterogeneous inputs. This study reveals us the complexities involved in the problem especially through the lack of standard Indian datasets which are compressed by the renowned legal databases manupatra, indiankanoon, etc.

We also found that DL models outperformed the existing feature-based models but are poor in reasoning the predictions made which can be overcome by attention models. Although the attention scores highlight the text regions that impact the predictions, they could not reach the justifications trusted by the legal practitioners. Mainly machine predictions will be credible based on the reason or justification provided which is an emerging concept to be addressed and needs attention.

REFERENCES

[1] George, A. (2017) 'The Problem of Pending Cases in Indian Courts: How to tackle?', ClearIAS, 9 April. Available at: https://www.clearias.com/pending-cases-in-indian-courts/ (Accessed: 30 May 2022).

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

[3] G. Sukanya and J Priyadarshini, "A meta-analysis of attention models on legal judgment prediction system," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, 2021.

[4] H. Zhong, Z. Guo, C. Tu, C. Xiao, Z. Liu, and M. Sun, "Legal judgment prediction via topological learning," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

[5] K. Branting et al. "Semi-supervised methods for explainable legal prediction," in Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, 2019.

[6] S. Long, C. Tu, Z. Liu, and M. Sun, "Automatic judgment prediction via legal reading comprehension," in Lecture Notes in Computer Science, Cham: Springer International Publishing, pp. 558–572, 2019.

[7] X. Guo, H. Zhang, L. Ye, and S. Li, "TenLa: an approach based on controllable tensor decomposition and optimized lasso regression for judgment prediction of legal cases," Appl. Intell., vol. 51, no. 4, pp. 2233–2252, 2021.

[8] Y. S. Chen, S.-W. Chiang, and M.-L. Wu, "A few-shot transfer learning approach using text-label embedding with legal attributes for law article prediction," Appl. Intell., vol. 52, no. 3, pp. 2884–2902, 2022.

[9] L. Li, L. Zhao, P. Nai, and X. Tao, "Charge prediction modeling with interpretation enhancement driven by double-layer criminal system," World Wide Web, vol. 25, no. 1, pp. 381–400, 2022.

[10] Bansal., A. Sharma, and R. K. Singh, "A review on the application of deep learning in the legal domain," in IFIP Advances in Information and Communication Technology, Cham: Springer International Publishing, pp. 374–381, 2019.

[11] X. Guo, H. Zhang, L. Ye, S. Li, and G. Zhang, "TenRR: An approach based on innovative tensor decomposition and optimized ridge regression for judgment prediction of legal cases," IEEE Access, vol. 8, pp. 167914–167929, 2020.

[12] V. G. Pillai and L. R. Chandran, "Verdict prediction for Indian courts using a bag of words and convolutional neural network," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020.

[13] V. Malik et al. "ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021.

[14] M. M. Janeela Theresa, and V. Joseph Raj, "A maximum spanning tree-based dynamic fuzzy supervised neural network architecture for classification of murder cases," Soft Comput., vol. 20, no. 6, pp. 2353–2365, 2016.

[15] 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.

[16] A.Thomas and S. Sangeetha, "Semi-supervised, knowledge-integrated pattern learning approach for fact extraction from judicial text," Expert Syst., vol. 38, no. 3, 2021.

[17] B. Chen, Y. Li, S. Zhang, H. Lian, and T. He, "A deep learning method for judicial decision support," in 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2019.

[18] C. Jin, G. Zhang, M. Wu, S. Zhou, and T. Fu, "Textual content prediction via fuzzy attention neural network model without predefined knowledge," China Commun., vol. 17, no. 6, pp. 211–222, 2020.

[19] C. Xiao, X. Hu, Z. Liu, C. Tu, and M. Sun, "Lawformer: A pre-trained language model for Chinese legal long documents," AI Open, vol. 2, pp. 79–84, 2021.

[20] N. Xu, P. Wang, L. Chen, L. Pan, X. Wang, and J. Zhao, "Distinguish confusing law articles for legal judgment prediction," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020.

[21] D. Li, Q. Zhao, J. Chen, and D. Zhao, "ADAN: An intelligent approach based on attentive neural network and relevant law articles for charge prediction," IEEE Access, vol. 9, pp. 90203–90211, 2021.

[22] L. Yue et al. "NeurJudge: A circumstance-aware neural framework for legal judgment prediction," in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval., 2021.

[23] Y. Huang, X. Shen, C. Li, J. Ge, and B. Luo, "Dependency learning for Legal Judgment Prediction with a unified text-to-text transformer," 2021.

[24] Yao, F., Sun, X., Yu H., Zhang W., Fu K., Commonalities Specificities, and Dependencies-Enhanced Multi-Task Learning Network for Judicial Decision Prediction. Neurocomputing (2020).

[25] F. Yao, X. Sun, H. Yu, Y. Yang, W. Zhang, and K. Fu, "Gated hierarchical multi-task learning network for judicial decision prediction," Neurocomputing, vol. 411, pp. 313–326, 2020.

[26] Y. Lyu et al. "Improving legal judgment prediction through reinforced criminal element extraction," Inf. Process. Manag., vol. 59, no. 1, p. 102780, 2022.

[27] L. K. Branting et al. "Scalable and explainable legal prediction," Artif. Intell. Law, vol. 29, no. 2, pp. 213–238, 2021.

[28] J. Li, G. Zhang, L. Yu, and T. Meng, "Research and design on cognitive computing framework for predicting judicial decisions," J. Signal Process. Syst., vol. 91, no. 10, pp. 1159–1167, 2019.

[29] X. Zhou, Y. Zhang, X. Liu, C. Sun, and L. Si, 'Legal intelligence for E-commerce: Multi-task learning by leveraging multiview dispute representation', in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019.

[30] H. Zhong, Y. Wang, C. Tu, T. Zhang, Z. Liu, and M. Sun, "Iteratively questioning and answering for interpretable legal Judgment Prediction," Proc. Conf. AAAI Artif. Intell., vol. 34, no. 01, pp. 1250–1257, 2020.

[31] L. Ma et al. "Legal judgment prediction with multi-stage case representation learning in the real court setting," in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval., 2021.

[32] 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," Inf. Syst., vol. 106, no. 101718, p. 101718, 2022.

[33] W. Hwang, D. Lee, K. Cho, H. Lee, and M. Seo, 'A multi-task benchmark for Korean legal language understanding and judgment prediction', 2022.

[34] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun, "How does NLP benefit legal system: A summary of legal artificial intelligence," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5218–5230, 2020.

[35] https:// manupatrafast.com

[36] https:// indiankanoon.org