**IAES International Journal of Artificial Intelligence (IJ-AI)**
Vol. 10, No. 4, December 2021, pp. 1091~1102
ISSN: 2252-8938, DOI: 10.11591/ijai.v10.i4.pp1091-1102 

# A systematic literature review of machine learning methods in predicting court decisions

**Nur Aqilah Khadijah Rosili[1], Noor Hidayah Zakaria[2], Rohayanti Hassan[3], Shahreen Kasim[4], Farid Zamani Che Rose[5], Tole Sutikno[6]**
[1,2,3]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
[4]Faculty of Computer Science and Information System, Universiti Tun Hussein Onn, Johor, Malaysia
[5]School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia
[6]Department of Electical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

**Article Info**

***Article history:***

Received Jun 12, 2021
Revised Sep 30, 2021
Accepted Oct 13, 2021

***Keywords:***

Judicial cases
Legal system
Machine learning
Predicting court decision
Predictive model
ROSES

**ABSTRACT**

Envisaging legal cases' outcomes can assist the judicial decision-making process. Prediction is possible in various cases, such as predicting the outcome of construction litigation, crime-related cases, parental rights, worker types, divorces, and tax law. the machine learning methods can function as support decision tools in the legal system with artificial intelligence's advancement. This study aimed to impart a systematic literature review (SLR) of studies concerning the prediction of court decisions via machine learning methods. The review determines and analyses the machine learning methods used in predicting court decisions. This review utilised RepOrting Standards for Systematic Evidence Syntheses (ROSES) publication standard. Subsequently, 22 relevant studies that most commonly predicted the judgement results involving binary classification were chosen from significant databases: Scopus and Web of Sciences. According to the SLR's outcomes, various machine learning methods can be used in predicting court decisions. Additionally, the performance is acceptable since most methods achieved more than 70% accuracy. Nevertheless, improvements can be made on the types of judicial decisions predicted using the existing machine learning methods.



***Corresponding Author:***

Nur Aqilah Khadijah Rosili
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia
Sultan Ibrahim Chancellery Building, Jalan Iman, 81310 Skudai, Johor, Malaysia
Email: aqilahrosili@gmail.com

## 1. INTRODUCTION

The globalised world today demands speedy and efficient handling of every action [1]–[3]. The fast-moving actions are essential in ensuring that the services can be implemented in line with the rapid development of technology and information, including in the legal system [4]–[20]. Judges and lawyers generally handle legal cases, but the help of technology is critically essential due to the massive numbers of cases daily. The effect of 'delay in justice' may lead to various consequences, such as witness hostility, unfitness of the plaintiff or accused and other adverse impacts [21].

Legal professionals currently focus on artificial intelligence [22]. According to historical datasets in the legal context, judicial decisions' prediction is standard and widely practised in the worldwide legal system. Machine learning is a budding scientific algorithms study, and statistical models are artificial intelligence's parts that enable systems to automatically learn and improvise experience from the test data [23]–[30].

The legal system's advancement via the usage of the machine learning algorithm is crucial in reducing the workload of legal professions and saves the time to settle pending cases during the Covid-19 pandemic

[21], [31]–[33]. Therefore, this study aimed to investigate the existing machine learning method developed to predict judicial decisions. the cases that used this approach were identified, and the methods' performance was monitored to study the methods' effectiveness.

## 2. METHOD

### 2.1. The Review Protocol-ROSES

The ROSES review protocol lead the current research. The ROSES protocol is developed for systematic review and environment management field maps [34]–[45]. Additionally, the ROSES protocol also encourages researchers to guarantee that they offer the correct information with explicit details. The researchers began the SLR by formulating research questions according to the review's protocol [46]–[48]. Subsequently, the researchers were required to describe the systematic searching strategy that consists of three processes, such as identification, screening and eligibility. the researchers were also required to perform a quality appraisal of the selected articles. Lastly, the authors elaborated on the outcomes generated from the chosen principal articles.

### 2.2. Formulation of Research Questions

The research questions for this study were formulated according to the elements of Population or Problem (P), Interest (I) and Context (Co), or PICo. The PICo is a tool to help researchers to construct research questions for the review. The PICo context encompasses the following aspects in this research: i) Population: Machine Learning, ii) Interest: Prediction, and iii) Context: Judicial Decision. The formulated research questions were:

1). What types of judicial decisions have been predicted using the machine learning method?
2). What are the machine learning methods used to predict judicial decisions?
3). How was the performance of the machine learning method used to predict judicial decisions?

### 2.3. Systematic Searching Strategies

The searching process in SLR comprises three main steps: i) identification, ii) screening, and iii) eligibility [7]. The whole process was summarised in the flow diagram depicted in Figure 1, and explained in the below sections.

#### 2.3.1. Identification

The purpose of the identification process is to maximise the number of keywords to be searched in databases. The keywords are developed based on the research questions formulated. The variation of keywords relies on an online thesaurus to identify synonyms and related terms, keywords used in previous studies and suggested by databases and experts. Nevertheless, the main keywords used in this study are prediction, judicial decision and machine learning. This study refers to two major indexed databases, namely Scopus and Web of Science. These databases were chosen due to several advantages.

First, the databases control the article's quality and consist of articles from various multidisciplinary fields. Second, the databases provide comprehensive and advance searching functions. The researchers constructed a full search string using the Boolean operator "AND" and "OR", phrase searching, truncation and wild card provided in both databases, as Table 1. Furthermore, the identification process also included manual searching to identify relevant articles in predicting judicial decisions using machine learning. This process managed to retrieve 94 articles from Scopus and 32 articles from Web of Science.

#### 2.3.2. Screening

The screening process was undertaken for all the selected articles in the identification process. The purpose of the screening purpose is to include and exclude articles based on the criteria determined. the initial screening process restricts the timeline to be in a specific interval recommended by Okoli [49]. The searching process was limited to articles published from the year 2000 to 2021 only. Nevertheless, the searching process was started in March 2021, and the year has not come to an end. Thus, the findings were limited to March 2021. the second inclusion criterion was the language used in the published articles or journals. All non-English language articles were excluded due to possible translation difficulties. The inclusion and exclusion criteria are enlisted in Table 2.

#### 2.3.3. Eligibility

The final process in the systematic searching procedure is eligibility. This process was undertaken manually to review the articles by reading all the articles' titles and abstracts thoroughly. The eligibility step

ensures that all the selected articles complied with the pre-determined criteria. the eligibility process included 20 articles retrieved from Scopus and 14 articles from Web of Science after manually reviewed.

Table 1. The search strings

| Database | Search String |
|---|---|
| Scopus | TITLE-ABS-KEY(("predict*" OR "prediction*" OR "predicting*" OR "forecast*") AND ("court decision*" OR "legal decision*" OR "law decision*" OR "judicial case*") AND ("machine learning*" OR "artificial intelligence*" OR "AI*" OR "supervise* machine learning*")) |
| Web of Science | (TS = (("prediction*" OR "predict*" OR "predicting") AND ("court decision*" OR "judicial decision*" OR "legal decision*") AND ("machine learning" OR "AI"))) |

Table 2. The inclusion and exclusion criteria

| Criteria | Inclusion | Exclusion |
|---|---|---|
| Timeline | 2000-2021 | Before 2000 |
| Language | English | Non-English |
| Methods | Machine learning | Other than machine learning |

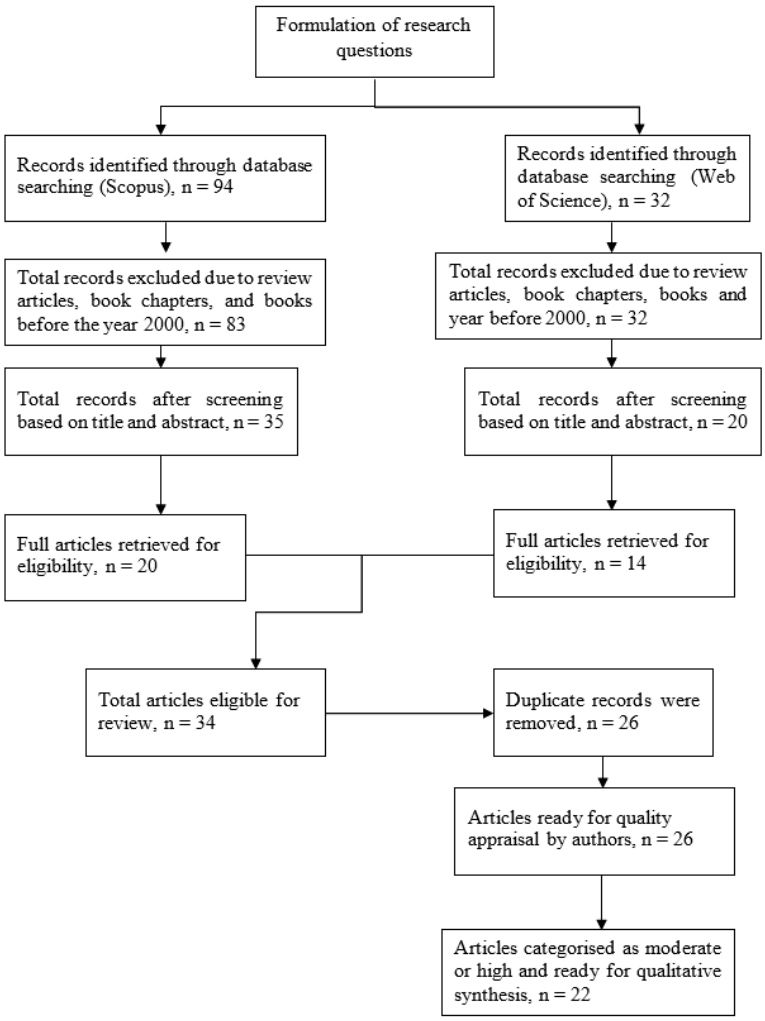


Figure 1. The flow diagram [50]

### 2.4. Quality Appraisal

The purpose of constructing Quality Assessment (QA) is to decide concerning the chosen studies' overall quality [22]. Thus, the following quality criteria were utilised to evaluate the chosen studies to figure out the strength of the studies' findings:

**QA1.** Does the study relate to the research objectives?

**QA2.** Does the study mention the method or approach used in prediction?
**QA3.** Is the research methodology clearly explained?
**QA4.** Is the data collection method described?
**QA5.** Does the performance of the method used have been discussed?

The 26 selected studies were examined through the five QA questions to determine the researchers' confidence in the chosen studies' credibility. Two experts were invited to appraise the QA to determine the articles' content quality. the reviewer ranked the articles into three levels: low, moderate, and high, as suggested by [51]. The articles ranked as moderate and high were eligible for review in the following process. The researchers adapted the scoring strategy employed by [52] to assess the articles' quality. The scoring of the quality evaluation was structured as: i) 1 point represents 'Yes', ii) 0.5 point represents 'Partly', and iii) 0 point represents 'No'. The scoring point ranked the articles into three categories: i) zero (0) to two (2) points were considered as low, ii) two-point-five (2.5) to three (3) points were considered as moderate, and iii) three-point-five (3.5) to five (5) points were considered as high. Finally, only 22 articles were eligible for QA after scoring was undertaken.

## 3. RESULTS AND DISCUSSION

The outcomes of the chosen significant studies, visualisation of publication year and the outline of the QA findings are summarised in the following sections.

### 3.1 Selected Primary Studies

Twenty-two studies were chosen through the SLR to identify the types of legal judgement cases that employ the machine learning method to envisage the findings. Subsequently, the machine learning methods used are listed, and the performance of each method is discussed. Table 3 summarises the selected studies and consists of the studies' identity (ID), the publications' titles, the articles' authors and the articles' publication year.

Table 3. Summary of selected primary studies

| ID | Title | Authors | Year |
|---|---|---|---|
| S1 | Predicting the Outcome of Construction Litigation Using Boosted Decision Trees | David Arditi & Thaveeporn Pulket | 2005 |
| S2 | Predicting the Outcome of Construction Litigation Using Particle Swarm Optimisation | KW Chau | 2005 |
| S3 | Prediction of Construction Litigation Outcome - A CBR Approach | KW Chau | 2006 |
| S4 | Predicting the Outcome of Construction Litigation Using an Integrated AI Model | David Arditi & Thaveeporn Pulket | 2010 |
| S5 | Litigation Outcome Prediction of Differing Site Condition Disputes Through Machine Learning Models | Tarek Mahfouz & Amr Kandil | 2012 |
| S6 | Study of Termination of Parental Rights: An Analysis of Israeli Court Decisions in Favour or Against Termination of Parental Rights | Vered Ben-David | 2016 |
| S7 | Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective | Nikolaos Aletras *et al* | 2016 |
| S8 | Learning to Predict Charges for Criminal Cases with Legal Basis | Bingfeng Luo *et al* | 2017 |
| S9 | A General Approach for Predicting the Behaviour of the Supreme Court of the United States | Daniel Martin Katz *et al* | 2017 |
| S10 | Predicting the Outcome of Appeal Decisions in Germany's Tax Law | Bernhard Waltl *et al* | 2017 |
| S11 | Legal judgement prediction via topological learning | Haoxi zhong *et al* | 2018 |
| S12 | Research and Design on Cognitive Computing Framework for Predicting Judicial Decision | Jiajing Li *et al* | 2018 |
| S13 | Using Machine Learning to Predict Decisions of the ECHR | Masha Medvedeva *et al* | 2019 |
| S14 | Predicting Outcomes of Legal Cases Based on Legal Factors Using Classifier | Rafe Athar Sheikh *et al* | 2019 |
| S15 | Predicting Disengagement from Judicial Proceedings by Female Victims of Intimate Partner Violence in Spain: A Systematic Replication with Prospective Data | Maria García-Jiménez *et al* | 2019 |
| S16 | Predicting Outcomes of Judicial Cases and Analysis Using Machine Learning | Prof Priyanka Bhilare *et al* | 2019 |
| S17 | A Deep Learning Method for Judicial Decision Support | Baogui Chen *et al* | 2019 |
| S18 | Determining Worker Type from Legal Text Data Using Machine Learning | Yifei Yin *et al* | 2020 |
| S19 | Deep Learning Algorithm for Judicial Judgement Prediction Based on BERT | Yongjun Wang *et al* | 2020 |
| S20 | Predicting the Litigation Outcome of PPP Project Disputes Between Public Authority and Private Partner Using an Ensemble Model | Xiaoxiao Zheng *et al* | 2020 |
| S21 | A Novel Approach on Argument Based Legal Prediction Model Using Machine Learning | Riya Sil *et al* | 2020 |
| S22 | Two-Layered Fuzzy Logic Based Model for Predicting Court Decisions in Construction Contract Disputes | Navid Bagherian-Marandi *et al* | 2021 |

### 3.2 Publication Years

The chosen studies were published between 2000 and 2021. Nevertheless, the earliest study published on this topic was from 2005. Figure 2 displays the number of studies published within the selected timeline. Nevertheless, the graph is not plotted for the year 2021, as the research for the particular year is still ongoing. Overall, the only latest study was published in January 2021, while four articles were published in 2020. Five articles were published in 2019, two in 2018, three in 2017 and two in 2016. Only one article was published in 2012, 2010, and 2006, whereas two articles were published in 2005. Based on the results, many studies were observed to have been published in the last five years. Therefore, the machine learning method can function as one of the approaches in improving the legal system by predicting outcomes.

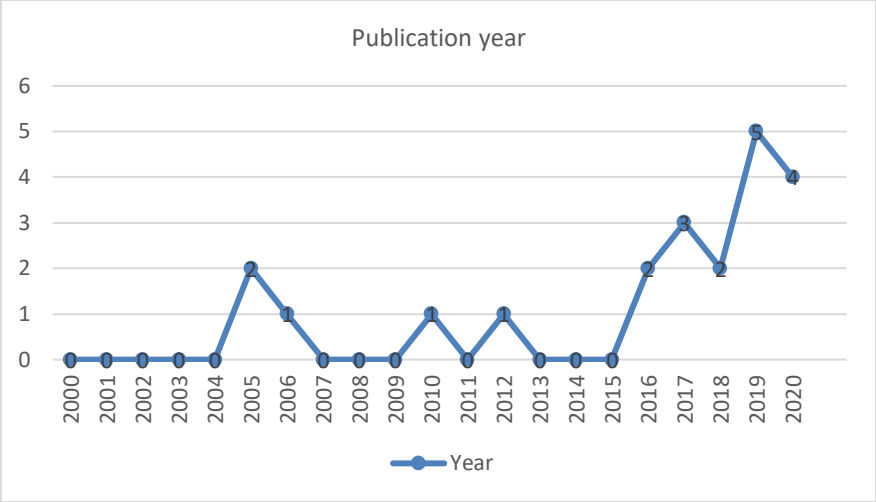


Figure 2. Number of selected studies over the years

### 3.3 QA Result

The chosen studies were assessed based on the QA questions explained in Section 2.4, and the analysis is presented in Table 4. The table demonstrates that 17 studies received high scores between the total score of three-point-five (3.5) to five (5), whereas five studies obtained a moderate score of 3. Conversely, four studies that obtained low scores were excluded from the review.

Table 4. QA results

| Study ID | QA1 | QA2 | QA3 | QA4 | QA5 | Score | Rating |
|---|---|---|---|---|---|---|---|
| S1 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 3 | Moderate |
| S2 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 3 | Moderate |
| S3 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 3 | Moderate |
| S4 | 1 | 1 | 1 | 1 | 1 | 5 | High |
| S5 | 1 | 1 | 0.5 | 1 | 0.5 | 4 | High |
| S6 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 3 | Moderate |
| S7 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 3 | Moderate |
| S8 | 1 | 1 | 1 | 1 | 1 | 5 | High |
| S9 | 1 | 1 | 1 | 1 | 1 | 5 | High |
| S10 | 1 | 1 | 0.5 | 0.5 | 1 | 4 | High |
| S11 | 1 | 1 | 0.5 | 1 | 1 | 4.5 | High |
| S12 | 1 | 1 | 1 | 1 | 1 | 5 | High |
| S13 | 1 | 1 | 1 | 1 | 1 | 5 | High |
| S14 | 1 | 1 | 0.5 | 0.5 | 1 | 4 | High |
| S15 | 1 | 1 | 0.5 | 0.5 | 0.5 | 3.5 | High |
| S16 | 1 | 1 | 0.5 | 1 | 1 | 4.5 | High |
| S17 | 1 | 1 | 1 | 1 | 0.5 | 4.5 | High |
| S18 | 1 | 1 | 1 | 1 | 1 | 5 | High |
| S19 | 1 | 1 | 1 | 0.5 | 1 | 4.5 | High |
| S20 | 1 | 1 | 0.5 | 1 | 1 | 4.5 | High |
| S21 | 1 | 1 | 0.5 | 1 | 1 | 4.5 | High |
| S22 | 1 | 1 | 1 | 1 | 1 | 5 | High |

These below sections provide a breakdown of the results according to the research questions identified in Section 2.2. The description of the findings from each research question is presented in separate subsections. the research questions are abbreviated as RQ hereafter.

### 3.4. Types of Judicial Decision

The research questions are discussed in this section. The first research question that was addressed: (RQ1) What types of judicial decisions have been predicted using the machine learning method? In the world of the legal system, judgement consists of various subtasks that have to be considered. The legal system is difficult to be understood by the civilians as the legal processes include interacting with a lawyer, hiring the lawyer, proceeding decisions and the legal decisions' consequences and the implications of words in the case files [53]. This study investigated how machine learning can be used in court proceedings to predict judicial decisions. the prediction can be of various types, such as predicting the legal judgement's outcome or the charges that require multilabel text classification. Multiple subtasks in legal judgement typically comprise comprehensive and complex sub-clauses, such as charges, penalty terms, and fines [52]. Nevertheless, most research experimented with a binary task that classifies only two possible outcomes. Besides predicting the outcome of judicial decision, several countries that utilise the civil law system, such as Germany, France and China, deemed that the prediction of relevant articles is a fundamental subtask that guides and supports the prediction [52].

In this SLR, seven research papers were found to have discussed envisaging construction litigation's outcome. Arditi and Phulket [54] mentioned that construction litigation is ordinary in numerous construction projects, explicitly involving large contracts. Miscommunication, insufficient specifications and plans, rigid contracts, changes in site conditions, non-payment, catch up profits, limited workforce, insufficient tools and equipment, ineffective supervision, notice requirements, constructive changes not acknowledged by owner, delays, and acceleration measures provoking claims and causing disputes. Therefore, Arditi and Phulket [54] proposed a tool to predict the outcome of litigation to minimise construction disputes caused by disagreements that are complicated to be settled without engaging in legal actions [54], [55].

Legal action requires a higher settlement cost because the litigation process is costly as the process involves complex issues. Additionally, the disagreement between client and contractor may lead to reputation damage on both sides [54]. In addition, legal action is time-consuming for complex construction disputes and may take two to six years before trial, depending on the jurisdiction [56]. Therefore, the researchers recommend several machine learning methods to ensure the accuracy of predicting a dispute resolution's outcome in courts. the methods can efficiently decrease the number of disputes that require higher spending costs through the litigation process [51].

According to the current study's findings, nine research papers predicted the outcome for crime-related cases. Nevertheless, crime-related cases can be divided into few categories. Aletras presented the first systematic study that predicted the outcome of cases in the European Court of Human Rights based on textual analysis [57]. The authors classified the prediction outputs into 'violation' and 'non-violation' based on text extracted from previous cases. Further studies were conducted by improving the number of articles and different variables using the same dataset [58]. This proposal can benefit lawyers and judges as a supporting tool to identify cases and extract text that guides decision-making [57].

Luo [59] asserted that the technique of analysing textual fact is crucial for legal assistant systems where civilians unfamiliar with legal terms can find similar cases or possible penalties by describing a case with their own words and understand the legal basis of their search cases. Furthermore, Luo [59] proposed an attention-based neural network method as a standard method to predict charges and extract relevant articles in a unified framework. The findings demonstrated that providing related articles can enhance the charge prediction results and envisage charges for cases with diverse expression styles effectively.

Zhong *et. al.* [60] proposed a different approach in modelling the judgement prediction framework that utilises multiple subtasks by claiming that previous studies only designed approaches for particular subtasks set and difficult to scale to other subtasks although developed to predict law articles and charges simultaneously. Additionally, it focused on murder related cases by undertaking such analysis. Extraction of legal judgement can be utilised to identify the details of case-specific legal factors but does not involve easy work and is time-consuming. Therefore, essential factors that will affect the prediction for murder related cases are evaluated by preparing a dataset to determine the factors as descriptors for prediction outcomes. The outcome prediction is viewed as a binary classification for classes as 'acquittal' and 'conviction' of the accused person.

The current study's finding is further discussed with cases that do not involve civil law and specifically focus on family law cases. Among the highlighted cases are disengagement, divorce, parental rights and dowry. Ben-David [61] conducted a crucial study regarding court decisions in 'favour' or 'against' the termination of parental rights that found the balance between the child's best interest, the parent's right and the privacy of the

family unit [62]. Li *et. al.* [63] proposed a prediction model for divorce. the research objectives were to predict the decisions for divorce cases with diverse expression styles and provide an easy understanding to the public regarding the results [60].

In addition, García-Jiménez *et. al.* [34] studied disengagement prediction where the researchers examined the variable needed by victims from legal proceedings before modelling the prediction model. This study developed a binary logistic regression model that predicts disengagement with two variables that are different from previous approaches. the first variable is the contact with the abuser, whereas the second variable is the interaction between the contact and thought of reuniting with the abuser. The paper aimed to predict disengagement by protecting women from being oppressed by court decisions. They believed that other factors should not influence court decisions in disengagement cases, such as not granted a protection order, not feeling supported by lawyers or unconvincing responses from professionals during proceedings [64].

Beneficiaries in India spent a long time waiting to get decisions from the court due to the scarcity of skilled workforce and infrastructure [21]. The prolonged legal proceeding may lead to various consequences. Sil *et al*. proposed a model that will assist legal professionals in analysing and performing predictions to give an outcome as 'guilty' or 'not guilty' depending on the parameters of death-related dowry cases [21]. A worker type approach has also been proposed in predicting court decisions for employment rights and protection purposes [65]. The outcome of various types of cases has been explored in predicting the outcome of court decisions using machine learning, leading to a conclusion that there are still opportunities and room for other cases to adapt the machine learning method as a supporting tool in decision-making. Future studies can include an extensive study on cases that require machine learning as a prediction model to lessen decision-making time.

### 3.5. Methods of Machine Learning

In this section, the following research question is discussed: (RQ2) What are the machine learning methods used to predict judicial decisions? Legal professionals are currently focused on artificial intelligence [66]. Envisaging judicial decisions based on historical datasets in the legal domain is not new and widely used in the legal system globally. Machine learning is an emerging scientific study of algorithms and statistical models that are part of artificial intelligence, enabling the system to learn automatically and improve the experience from test data. The core research aspects in applying machine learning in jurisprudence are the extraction of information and analysis on existing legal documents. in previous practices, lawyers and judges have to do all the works manually. However, machine learning has taken the stream of society to become more intelligent by interpreting the text documents and extracting the documents' content [53].

The researchers observed the proposed machine learning in this SLR by determining the types and names of the classifier used in predicting judicial decisions. the majority of studies attempted to extricate efficient features from text content or case annotations (dates, terms, locations, and types) [1]. Nevertheless, Zhong *et al.* [60] asserted that the conventional methods could only employ shallow textual features and manually designed factors. the features and factors need enormous human efforts and regularly undergo generalisation problems when applied in other scenarios. the achievement of neural networks on natural language processing (NLP) tasks inspired the researchers to start handling legal judge prediction by integrating neural models with legal knowledge [59]. Luo [59] laid out an attention-based neural network that jointly models charge prediction and relevant article extraction. Nonetheless, these models are designed for specific subtasks. Therefore, non-trivial elements should be widened to other subtasks of legal judge prediction with complex dependencies.

The current study researchers classified the methods using two types: single classifier and combined classifier. Subsequently, the researchers identified the name of the classifier(s) involved as the prediction model. the single classifier refers to an individual model of machine learning that is used in the prediction. In contrast, combined classifiers refer to an ensemble model that used more than one classifier in making predictions. As shown in Table 5, the most common classifier is the support vector machine (SVM). According to the current SLR, six papers proposed SVM as the prediction model in various cases.

Nevertheless, this finding cannot be concluded as the preferred method in prediction as other models also displayed a good performance in predicting judicial decisions depending on cases. The ensemble method provides an enhanced approach when compared with another approach. Thus, the researchers concluded that this research area is still new and open for exploration. This research is still actively ongoing in the recent five years, as observed in Figure 2. Therefore, a great opportunity is present for further research concerning implementing machine learning methods in predicting court decisions.

Table 5. Summary of methods used

| Types of cases | Method | Classifier(s) involved | Study ID |
|---|---|---|---|
| Construction Litigation | Single Classifier | Boosted Decision Tree (BDT) | S1 |
| | | Particle Swarm Optimisation (PSO) | S2 |
| | | Case Base Reasoning (CBR) | S3 |
| | | Integrated Prediction Model (IPM) | S4 |
| | | Support Vector Machine (SVM) | S5 |
| | | Two-Layered Fuzzy Logic | S22 |
| | Combined Classifier | Gradient Boosting Decision Tree (GBDT), k-nearest neighbour (KNN), Multilayer Perceptron (MLP) | S20 |
| Crime | Single Classifier | Support Vector Machine (SVM) | S7, S8, S13, S16 |
| | | Random Forest (RF) | S9 |
| | | Multi Task Learning (MTL) | S11 |
| | | Classification and Regression Trees (CART) | S14 |
| | | Convolutional Neural Network (CNN) | S17 |
| | Combined Classifier | Bidirectional Encoder Representation from Transformer (BERT) + Convolutional Neural Network (CNN) | S19 |
| Worker type | Single Classifier | extended Multilayer Perceptron (eMLP) | S18 |
| Disengagement | | Logistic Regression (LR) | S15 |
| Tax Law | | Naïve Bayes (NB) | S10 |
| Parental Right | | Logistic Regression (LR) | S6 |
| Divorce | | Cognitive Computing Framework (CCF) | S12 |
| Dowry | | Support Vector Machine (SVM) | S21 |

### 3.6. Performance of the Machine Learning Methods

The following research question is addressed in this section: (RQ3) How was the performance of machine learning methods used to predict judicial decisions? The performance of the prediction model proposed should be assessed prior to understanding the approach used. the efficiency of any machine learning model can be measured through k-fold cross-validation, accuracy, sensitivity, specificity, recall, precision, and F-measure [63]. Based on the observations from the 22 reviewed papers, most researchers used accuracy, precision, recall and F-measure in evaluating the performance of their models. F-measure, precision and recall are frequently utilised in extracting information as performance measurement since machine learning performance assessments include specific trade-off levels between true positive and true negative rates [63].

Table 6 summarises the information regarding accuracy, precision, recall or sensitivity adapted from [21]. There are four important terms used in measuring the performance metrics, namely true positive (tp), true negative (tn), false positive (fp) and false negative (fn) [21]. Earlier research (S1, S2, S3 and S4) used different approaches in evaluating the performance of the methods used. the average prediction rate generated in the reported study is within the range of 80% to 91%. Nevertheless, the study was expanded into the next stage by adjusting the number and format of attributes and the number of cases used to better predict rates [67].

Table 6. Performance metrics formula

| Measure of Performance | Description | Formula |
|---|---|---|
| Accuracy | The ratio of a correctly predicted result to the total actual result | $\frac{tp + tn}{tp + tn + fp + fn}$ |
| Precision | The ratio of a correctly predicted positive result to the total positive predicted result | $\frac{tp}{tp + fp}$ |
| Recall of Sensitivity | The ratio of a correctly predicted positive result to the total result | $\frac{tp}{tp + fn}$ |
| F1 Score | The weighted average of precision and recall if the class distribution is uneven | $\frac{2(recall * precision)}{recall + precision}$ |

The most intriguing finding of the SLR found is that 16 out of the 22 selected review papers obtained more than 80% of accuracy, precision or prediction rate through the evaluation process. Only four papers (S7, S10 S15 and S22) obtained the range of accuracy or precision of 50% to 70%. Conversely, two papers (S6 and S13) did not discuss the performance of their prediction models in detail. the summary of the performance results of the 22 reviewed papers is presented in Table 7. This approach explicitly observed that the prediction model could be a reliable supporting tool in determining court decisions as the models' performance achieved more than 70% overall accuracy.

Table 7. Results of performance

| Study ID | Model | Results | | | | |
|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Recall | F1 Score | Prediction rate |
| S1 | BDT | | | | | 90 |
| S2 | PSO | | | | | 80 |
| S3 | CBR | | | | | 84 |
| S4 | IPM | | | | | 91 |
| S5 | SVM | 98 | 98 | 98 | 98 | |
| S6 | LR | | | | | |
| S7 | SVM | 79 | | | | |
| S8 | SVM | | 98 | 95 | 97 | |
| S9 | RF | | 70 | 70 | 69 | |
| S10 | NB | | 57 | 57 | 57 | |
| S11 | MTL | 95.6 | 75.9 | 69.6 | 70.9 | |
| S12 | CCF | | 71.22 | 74.17 | 72.65 | |
| S13 | SVM | | | | | |
| S14 | CART | 91.86 | 92.86 | 90.7 | 91.76 | |
| S15 | LR | 74.7 | 74.4 | 76.2 | | |
| S16 | SVM | | 92 | 91 | 91 | |
| S17 | CNN | | 88.75 | | 86.27 | |
| S18 | eMLP | 91.7 | 89.4 | 90.6 | 90 | |
| S19 | BERT, CNN | | 89.7 | 89.7 | 89.6 | |
| S20 | GBDT, KNN, MLP | 96.42 | 96.66 | 96.38 | 96.03 | |
| S21 | SVM | 93 | 93 | 93 | 92 | |
| S22 | Two-layered Fuzzy Logic | 73.9 | | | | |

## 4. CONCLUSION

This study has presented an investigation regarding predicting court decisions using machine learning methods. The importance of predicting judicial decisions can be identified in various cases and from the research outcome obtained. This approach can improvise the legal system by making it more systematic and reliable. the methods and features derived from the findings could fill the existing gaps in the study area for future scholarly work. This systematic review study is expected to contribute to the body of knowledge by providing an overview regarding existing models used in predicting judicial decisions, the performance of the predicting model and discussion on several types of cases in the legal system that adapted this approach. The review also offers several recommendations for future studies, including new types of cases for predicting judicial decisions and a new machine learning method that requires a combined classifier to improve the predicting tools' performance.

## ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi Malaysia, Universiti Tun Hussein Onn, Universiti Sains Malaysia and Universitas Ahmad Dahlan to support this collaborative research.

## REFERENCES

[1] D. M. Katz, M. J. Bommarito, and J. Blackman, "A general approach for predicting the behavior of the Supreme Court of the United States," *PLoS One*, vol. 12, no. 4, pp. e0174698–e0174698, 2017.

[2] N. Kamaruddin, R. D. Safiyah, and A. Wahab, "Small and medium enterprise business solutions using data visualization," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6. pp. 2562–2568, 2020, doi: 10.11591/eei.v9i6.2463.

[3] J. Shetty, B. Sathish Babu, and G. Shobha, "Proactive cloud service assurance framework for fault remediation in cloud environment," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1. pp. 987–996, 2020, doi: 10.11591/ijece.v10i1.pp987-996.

[4] A. V Zadgaonkar and A. J. Agrawal, "An overview of information extraction techniques for legal document analysis and processing," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6. pp. 5450–5457, 2021, doi: 10.11591/ijece.v11i6.pp5450-5457.

[5] I. E. Olufemi, A. A. Adebiyi, F. A. Ibikunle, M. O. Adebiyi, and O. O. Oludayo, "Research trends on CAPTCHA: A systematic literature," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5. pp. 4300–4312, 2021, doi: 10.11591/ijece.v11i5.pp4300-4312.

[6] A. Murugan, T. Chechare, B. Muruganantham, and S. Ganesh Kumar, "Healthcare information exchange using blockchain technology," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1. pp. 421–426, 2020, doi: 10.11591/ijece.v10i1.pp421-426.

[7] V. Siddesh Padala, K. Gandhi, and D. V Pushpalatha, "Machine learning: The new language for applications," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 4. pp. 411–421, 2019, doi: 10.11591/ijai.v8.i4.pp411-421.

[8] A. M. Muad, N. S. M. Bahaman, A. Hussain, and M. Y. P. M. Yusof, "Tooth segmentation using dynamic

programming-gradient inverse coefficient of variation," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 1. pp. 253–260, 2019, doi: 10.11591/eei.v8i1.1446.
[9] G. Hima Bindu, C. Anuradha, and P. S. R. Chandra Murty, "A survey on multimedia content protection mechanisms," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6. pp. 4204–4211, 2018, doi: 10.11591/ijece.v8i6.pp.4204-4211.
[10] S. K. Srivastava, "Artificial intelligence: Way forward for India," *IAES International Journal of Artificial Intelligence*, vol. 7, no. 1. pp. 19–32, 2018, doi: 10.11591/ijai.v7.i1.pp19-32.
[11] N. Madhusudan and L. Manjunatha Rao, "Insights on extent of effectiveness, trend, and gap in existing frameworks for e-procurement system," *International Journal of Electrical and Computer Engineering*, vol. 6, no. 2. pp. 751–758, 2016, doi: 10.11591/ijece.v6i2.9028.
[12] O. J. Ayangbekun, O. F. Bankole, and B. A. Saka, "Analysis of security mechanisms in Nigeria E-banking platform," *International Journal of Electrical and Computer Engineering*, vol. 4, no. 6. pp. 837–847, 2014, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84928530364&partnerID=40&md5=fa9dcf0ac0f9e8933c4121ec469d9adf.
[13] Z. Anna and E. Vladimir, "State regulation of the IoT in the Russian Federation: Fundamentals and challenges," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5. pp. 4542–4549, 2021, doi: 10.11591/ijece.v11i5.pp4542-4549.
[14] P. Anki and A. Bustamam, "Measuring the accuracy of LSTM and BiLSTM models in the application of artificial intelligence by applying chatbot programme," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1. pp. 197–205, 2021, doi: 10.11591/ijeecs.v23.i1.pp197-205.
[15] I. Odun-Ayo, O. Alagbe, and J. Yahaya, "A systematic mapping study of security, trust and privacy in clouds," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3. pp. 1598–1610, 2021, doi: 10.11591/eei.v10i3.1887.
[16] N. Nizamuddin and A. Abugabah, "Blockchain for automotive: An insight towards the IPFS blockchain-based auto insurance sector," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3. pp. 2443–2456, 2021, doi: 10.11591/ijece.v11i3.pp2443-2456.
[17] I. Veiksa, "Protection of computer-generated works in the era of new technologies," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 1. pp. 234–243, 2021, doi: 10.11591/ijai.v10.i1.pp234-243.
[18] S. Nasir, S. M. Al-Qaraawi, and M. S. Croock, "QR based management system for plants shopping center," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 2. pp. 931–939, 2020, doi: 10.11591/ijeecs.v19.i2.pp931-939.
[19] V. Kasinathan, A. Mustapha, M. F. C. Abdul Rani, and S. A. Mostafa, "The role of chatterbots in enhancing tourism: A case study of Penang tourism spots," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4. pp. 569–575, 2020, doi: 10.11591/ijai.v9.i4.pp569-575.
[20] A. Zharova, "The protect mobile user data in Russia," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3. pp. 3184–3192, 2020, doi: 10.11591/ijece.v10i3.pp3184-3192.
[21] P. Bhilare, N. Parab, N. Soni, and B. Thakur, "Predicting outcome of judicial cases and analysis using machine learning," *Int. Res. J. Eng. Technol.*, vol. 6, no. 3, pp. 326–330, 2019.
[22] B. Kitchenham *et al.*, "Systematic literature reviews in software engineering--a tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010.
[23] K. J. Danjuma, "Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients," *arXiv Prepr. arXiv1504.04646*, 2015.
[24] Hertina *et al.*, "Data mining applied about polygamy using sentiment analysis on twitters in indonesian perception," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4. pp. 2231–2236, 2021, doi: 10.11591/EEI.V10I4.2325.
[25] A. Al Mamun, P. P. Em, and J. Hossen, "Lane marking detection using simple encode decode deep learning technique: SegNet," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4. pp. 3032–3039, 2021, doi: 10.11591/ijece.v11i4.pp3032-3039.
[26] R. Usha and K. Perumal, "A modified fractal texture image analysis based on grayscale morphology for multi-model views in MR Brain," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 1. pp. 154–163, 2021, doi: 10.11591/ijeecs.v21.i1.pp154-163.
[27] J. Y. W. Jien, A. Baharum, S. H. A. Wahab, N. Saad, M. Omar, and N. A. M. Noor, "Age-based facial recognition using convoluted neural network deep learning algorithm," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 3. pp. 424–428, 2020, doi: 10.11591/ijai.v9.i3.pp424-428.
[28] A. Al-Imam, M. A. Motyka, and M. Z. Jędrzejko, "Conflicting opinions in connection with digital superintelligence," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 2. pp. 336–348, 2020, doi: 10.11591/ijai.v9.i2.pp336-348.
[29] L. M. Shi, A. Mustapha, and Y. M. M. Hassim, "Predicting fatalities among shark attacks: Comparison of classifiers," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 4. pp. 360–366, 2019, doi: 10.11591/ijai.v8.i4.pp360-366.
[30] S. A. Diwan, "Proposed study on evaluating and forecasting the resident property value based on specific determinants by case base reasoning and artificial neural network approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 3. pp. 1467–1473, 2019, doi: 10.11591/ijeecs.v17.i3.pp1467-1473.
[31] H. A. Razak, M. A. M. Saleh, and N. M. Tahir, "Review on anomalous gait behavior detection using machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5. pp. 2090–2096, 2020, doi:

10.11591/eei.v9i5.2255.
[32] M. Ghosh and P. Prabu, "Empirical analysis of ensemble methods for the classification of robocalls in telecommunications," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4. pp. 3108–3114, 2019, doi: 10.11591/ijece.v9i4.pp3108-3114.
[33] B. Subeno, R. Kusumaningrum, and Farikhin, "Optimisation towards Latent Dirichlet Allocation: Its Topic Number and Collapsed Gibbs Sampling Inference Process," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 5. pp. 3204–3213, 2018, doi: 10.11591/ijece.v8i5.pp.3204-3213.
[34] M. García-Jiménez, M. J. Cala-Carrillo, and M. E. Trigo, "Predicting Disengagement from Judicial Proceedings by Female Victims of Intimate Partner Violence in Spain: A Systematic Replication With Prospective Data," *Violence Against Women*, vol. 26, no. 12–13, pp. 1493–1516, Oct. 2020, doi: 10.1177/1077801219882502.
[35] G. Rajalakshmi, A. Gopal, and R. Pandian, "An approach to assess the quality of honey using partial least square method," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6. pp. 4991–4998, 2021, doi: 10.11591/ijece.v11i6.pp4991-4998.
[36] S. Natarajan and R. S. R. Babu, "Comparison of cascaded H-bridge inverters for harmonic mitigation considering various loads," *International Journal of Power Electronics and Drive Systems*, vol. 8, no. 1. pp. 10–19, 2017, doi: 10.11591/ijpeds.v8.i1.pp10-19.
[37] S. R. Akbar, M. T. Handono, and A. Basuki, "Design of pervasive discovery, service and control for smart home appliances: An integration of Raspberry Pi, UPnP protocols and Xbee," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 2. pp. 1012–1022, 2017, doi: 10.11591/ijece.v7i2.pp1012-1022.
[38] M. Adebiyi and O. O. Olugbara, "Binding site identification of COVID-19 main protease 3D structure by homology modeling," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3. pp. 1713–1721, 2021, doi: 10.11591/ijeecs.v21.i3.pp1713-1721.
[39] F. 'A. Nor Rashid and N. S. Suriani, "Spiking neural network classification for spike train analysis of physiotherapy movements," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 1. pp. 319–325, 2020, doi: 10.11591/eei.v9i1.1868.
[40] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2. pp. 1015–1027, 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.
[41] M. Aljarah, M. Shurman, and S. H. Alnabelsi, "Cooperative hierarchical based edge-computing approach for resources allocation of distributed mobile and IoT applications," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1. pp. 296–307, 2020, doi: 10.11591/ijece.v10i1.pp296-307.
[42] S. A. A. Qadri, T. S. Gunawan, M. F. Alghifari, H. Mansor, M. Kartiwi, and Z. Janin, "A critical insight into multi-languages speech emotion databases," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4. pp. 1312–1323, 2019, doi: 10.11591/eei.v8i4.1645.
[43] N. A. N. Hashim, J. T. H. Loong, A. Ghazali, and F. A. Hamid, "Memristor based ring oscillators true random number generator with different window functions for applications in cryptography," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1. pp. 201–209, 2019, doi: 10.11591/ijeecs.v14.i1.pp201-209.
[44] W. Kurniawan, M. H. H. Ichsan, and S. R. Akbar, "UDP pervasive protocol implementation for smart home environment on MyRIO using LabVIEW," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 1. pp. 113–123, 2018, doi: 10.11591/ijece.v8i1.pp113-123.
[45] N. Suresh and R. S. R. Babu, "Reduction of total harmonic distortion in cascaded H-bridge inverter by pattern search technique," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 6. pp. 3292–3298, 2017, doi: 10.11591/ijece.v7i6.pp3292-3298.
[46] B. L. Putro, Y. Rosmansyah, and Suhardi, "An intelligent agent model for learning group development in the digital learning environment: A systematic literature review," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3. pp. 1159–1166, 2020, doi: 10.11591/eei.v9i3.2009.
[47] N. A. Ahmad, S. M. Drus, and N. A. A. Bakar, "Enterprise architecture adoption issues and challenges: A systematic literature review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1. pp. 399–408, 2019, doi: 10.11591/ijeecs.v15.i1.pp399-408.
[48] B. Jokonowo, J. Claes, R. Sarno, and S. Rochimah, "Process mining in supply chains: A systematic literature review," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6. pp. 4626–4636, 2018, doi: 10.11591/ijece.v8i6.pp4626-4636.
[49] C. Okoli, "A guide to conducting a standalone systematic literature review," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 1, p. 43, 2015.
[50] H. A. M. Shaffril, N. Ahmad, S. F. Samsuddin, A. A. Samah, and M. E. Hamdan, "Systematic literature review on adaptation towards climate change impacts among indigenous people in the Asia Pacific regions," *J. Clean. Prod.*, vol. 258, p. 120595, 2020.
[51] M. Petticrew and H. Roberts, *Systematic reviews in the social sciences: A practical guide*. John Wiley \& Sons, 2008.
[52] H. Alsolai and M. Roper, "A systematic literature review of machine learning techniques for software maintainability prediction," *Inf. Softw. Technol.*, vol. 119, p. 106214, 2020.
[53] N. Bagherian-Marandi, M. Ravanshadnia, and M.-R. Akbarzadeh-T, "Two-layered fuzzy logic-based model for predicting court decisions in construction contract disputes," *Artif. Intell. Law*, pp. 1–32, 2021.
[54] D. Arditi and T. Pulket, "Predicting the outcome of construction litigation using boosted decision trees," *J. Comput. Civ. Eng.*, vol. 19, no. 4, pp. 387–393, 2005.
[55] K. Nikolskaia and V. Naumov, "Artificial Intelligence in Law," in *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, 2020, pp. 1–4.

[56] R. Mothukuri and B. Rao, "Data Mining on Prediction of Crime and Legal Judgements: A State of an Art," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1, pp. 3670–3679, 2019.
[57] M. Medvedeva, M. Vols, and M. Wieling, "Using machine learning to predict decisions of the European Court of Human Rights," *Artif. Intell. Law*, vol. 28, no. 2, pp. 237–266, 2020.
[58] M. Gusenbauer and N. R. Haddaway, "Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources," *Res. Synth. Methods*, vol. 11, no. 2, pp. 181–217, 2020.
[59] B. Luo, Y. Feng, J. Xu, X. Zhang, and D. Zhao, "Learning to predict charges for criminal cases with legal basis," *arXiv Prepr. arXiv1707.09168*, 2017.
[60] 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, pp. 3540–3549.
[61] V. Ben-David, "Study of termination of parental rights: An analysis of Israeli court decisions in favor or against termination of parental rights," *J. Fam. Soc. Work*, vol. 18, no. 4, pp. 225–242, 2015.
[62] R. Sil and A. Roy, "A Novel Approach on Argument based Legal Prediction Model using Machine Learning," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, 2020, pp. 487–490.
[63] 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.
[64] K.-W. Chau, "Prediction of construction litigation outcome--a case-based reasoning approach," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2006, pp. 548–553.
[65] T. Mahfouz and A. Kandil, "Litigation outcome prediction of differing site condition disputes through machine learning models," *J. Comput. Civ. Eng.*, vol. 26, no. 3, pp. 298–308, 2012.
[66] Z. Liu and H. Chen, "A predictive performance comparison of machine learning models for judicial cases," in *2017 IEEE Symposium series on computational intelligence (SSCI)*, 2017, pp. 1–6.
[67] D. Arditi and T. Pulket, "Predicting the outcome of construction litigation using an integrated artificial intelligence model," *J. Comput. Civ. Eng.*, vol. 24, no. 1, pp. 73–80, 2010.