

A Systematic Literature Review of Machine Learning Methods in Predicting Court Decisions

By Nur Aqilah Khadijah Rosili

A Systematic Literature Review of Machine Learning Methods in Predicting Court Decisions

Nur Aqilah Khadijah Rosili¹, Rohayanti Hassan², Noor Hidayah Zakaria³, Shahreen Kasim⁴, Farid Zamani Che Rose⁵

^{1,2,3}School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM)

⁴Faculty of Computer Science and Information System, Universiti Tun Hussein Onn (UTHM)

⁵School of Mathematical Sciences, Univeristi Sains Malaysia (USM)

Article Info

Article history:

Received Jun 12, 201x
Revised Aug 20, 201x
Accepted Aug 26, 201x

Keywords:

Legal system
Machine learning
Judicial cases
Predictive model

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 22 sion tools in the legal system with artificial intelligence's advancement. This study aim 3 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 ROSES (RepOrting Standards for Systematic Evidence Syntheses) 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.

25
This is an open access article under the [CC BY-SA license](#).



Corresponding Author:

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

1. INTRODUCTION

The globalised world today demands speedy and efficient handling of 3 every action. 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. 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 [20].

Legal professionals currently focus on artificial intelligence [1]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 [11].

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 [20]. 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.

1

2. METHODOLOGY

2.1. The Review Protocol – ROSES

1 The ROSES review protocol lead the current research. The ROSES protocol is developed for systematic review and environment management field maps [14]. Additionally, the ROSES protocol also encourages 1 researchers to guarantee that they offer the correct information with explicit details. The researchers began the SLR by formulating research 1 questions according to the review's protocol. 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.

1

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:

Population: Machine Learning

Interest : Prediction

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: identification, screening, and eligibility [7]. The whole process was summarised in the flow diagram depicted in Figure 1.

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. 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.

Table 1. The search string

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*))	43
Web of Science	(TS = (("prediction*" OR "predict*" OR "predicting*") AND ("court decision*" OR "judicial decision*" OR "legal decision*") AND ("machine learning" OR "AI")))	

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 (1). 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 article (16) or journals. All non-English language articles were excluded due to possible translation difficulties. The inclusion and exclusion criteria are enlisted in Table 2 below.

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

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.

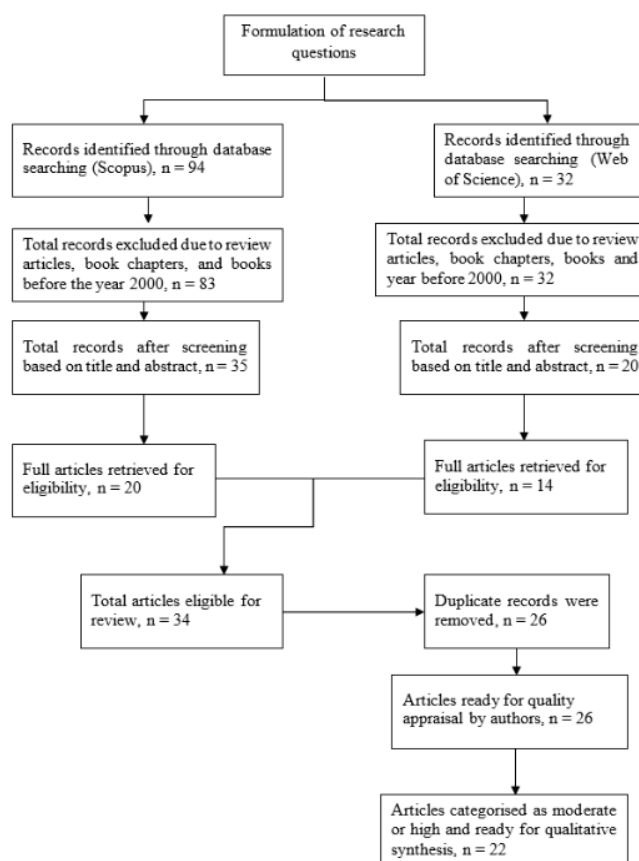


Figure 1. The flow diagram [7]

2.4. Quality Appraisal

35

The purpose of constructing Quality Assessment (QA) is to decide concerning the chosen studies' overall quality [1]. 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 Petticrew & Roberts(2006) The articles ranked as moderate and high were eligible for review in the following process. The researchers adapted the scoring strategy employed by Alsolai & Roper(2020) to assess the articles' quality. The scoring of the quality evaluation was structured as follows:

- 1 point represents 'Yes',
- 0.5 point represents 'Partly',
- 0 point represents 'No'.

The scoring point ranked the articles into three categories: Zero (0) to two (2) points were considered as low, two-point-five (2.5) to three (3) points were considered as moderate, and 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

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

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. This table consists of the studies' identity (ID), the publications' titles, the articles' authors and the articles' publication year.

Table 3. Summary of selected primary studies

Study 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 Your 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 Waltd et al	2017
S11	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	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 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.

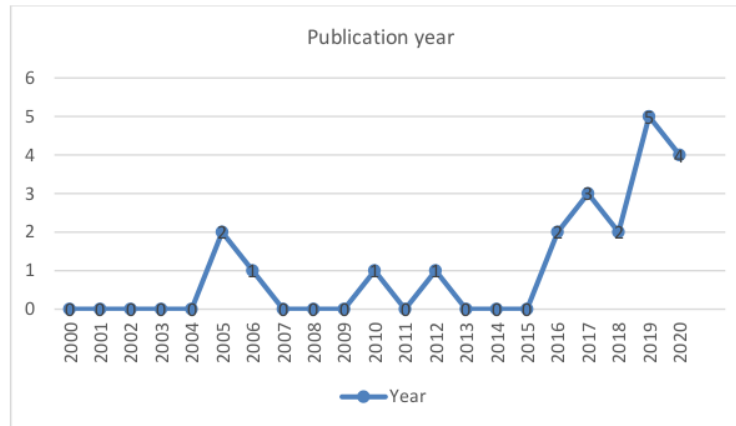


Figure 2. Number of selected studies over the years

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.

36
36

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

4. DISCUSSION

This section provides 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.

4.1. 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 [19]. 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 [8]. 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 [8].

In this SLR, seven research papers were found to have discussed envisaging construction litigation's outcome. Arditi and Phulket mentioned that construction litigation is ordinary in numerous construction projects, explicitly involving large contracts [4]. 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 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 [4, 12].

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 [4]. In addition, legal action is time-consuming for complex construction disputes and may take two to six years before trial, depending on the jurisdiction [21]. 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 [17].

According to the current study's findings, nine research papers predicted the outcome for 20 me-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 [16]. 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 [15]. This proposal can benefit lawyers and judges as a supporting tool to identify cases and extract text that guides decision-making [16].

Luo 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 [2]. Furthermore, Luo 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 [2].

Zhong et al. (2018) 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, Shaikh et al., (2020) 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: disengagement, divorce, parental rights and dowry. Bend David (2015) conducted a judicial study regarding court decisions in 'favour' or 'against' the termination of parental rights that find the balance between the child's best interest, the parent's right and the privacy of family unit [22]. Li et al., (2019) 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 [9].

In addition, García-Jiménez et al., (2020) studied disengagement prediction where the researchers examined the variable needed by victims from legal proceedings before modelling the prediction model. This study

17

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 [13].

Beneficiaries in India spent a long time waiting to get decisions from the court due to the scarcity of skilled workforce and infrastructure [20]. 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 [20]. A worker type approach has also been proposed in predicting court decisions for employment rights and protection purposes [23]. 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.

44

4.2. 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 Liu & Chen (2017) (Zhenyu, 2018). 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 (Ullah et al., 2020). 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 [19].

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 (e.g., dates, terms, locations, and types) [6]. Nevertheless, Zhong et al. (2018) 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 [2]. Luo laid out an attention-based neural network that jointly models charge prediction and relevant article extraction [2]. 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 (Refer to Table 5). 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
	8 Combined Classifier	Gradient Boosting Decision Tree (GBDT), k-nearest neighbour (KNN), Multilayer Perceptron (MLP)	S20
	Crime	Single Classifier	Support Vector Machine (SVM)
Random Forest (RF)			S9
Multi Task Learning (MTL)			S11
Classification and Regression Trees (CART)			S14
Convolutional Neural Network (CNN)			S17
19 Combined Classifier		Bidirectional Encoder Representation from Transformer (BERT) + Convolutional Neural Network (CNN)	S19
Worker type		Single Classifier	extended Multilayer Perceptron (eMLP)
Disengagement	Single Classifier	Logistic Regression (LR)	S15
Tax Law	Single Classifier	Naïve Bayes (NB)	S10
Parental Right	Single Classifier	Logistic Regression (LR)	S6
Divorce	Single Classifier	Cognitive Computing Framework (CCF)	S12
Dowry	Single Classifier	Support Vector Machine (SVM)	S21

4.3. 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 [10]. 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 [10].

Table 6 summarises the information regarding accuracy, precision, recall or sensitivity adapted from Bhilare et al., (2019) 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) [20]. 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 [5].

Table 6. Performance metrics formula

Measure of Performance	Description	Formula
Accuracy	16 The ratio of a correctly predicted result to the total actual result	7 $\frac{tp + tn}{tp + tn + fp + fn}$
Precision	37 The ratio of a correctly predicted positive result to the total positive predicted result	7 $\frac{tp}{tp + fp}$
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	$\frac{2(\text{recall} * \text{precision})}{\text{recall} + \text{precision}}$

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	Accuracy	Precision	Recall	F1 Score
S1	80%	70%	70%	73%
S2	80%	70%	70%	73%
S3	80%	70%	70%	73%
S4	80%	70%	70%	73%
S5	80%	70%	70%	73%
S6	50%	50%	50%	50%
S7	80%	70%	70%	73%
S8	80%	70%	70%	73%
S9	80%	70%	70%	73%
S10	80%	70%	70%	73%
S11	80%	70%	70%	73%
S12	80%	70%	70%	73%
S13	50%	50%	50%	50%
S14	80%	70%	70%	73%
S15	80%	70%	70%	73%
S16	80%	70%	70%	73%
S17	80%	70%	70%	73%
S18	80%	70%	70%	73%
S19	80%	70%	70%	73%
S20	80%	70%	70%	73%
S21	80%	70%	70%	73%
S22	80%	70%	70%	73%

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				

5. 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 improve 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.

REFERENCES

- [1] B. Kitchenham, et al., "Systematic Literature Reviews in Software Engineering-A Tertiary Study," *Information and Software Technology*, vol. 52, pp. 792-805, 2010.
- [2] B. Luo, et al., "Learning to Predict Charges for Criminal Cases with Legal Basis," in *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pp. 2727-2736, 2017.
- [3] C. Okoli, "A Guide to Conducting a Standalone Systematic Literature Review," *Communications of the Association for Information Systems*, vol. 37, pp. 879-910, 2015.
- [4] D. Arditi and T. Pulket, "Predicting the Outcome of Construction Litigation Using Boosted Decision Trees," *Journal of Computing in Civil Engineering*, vol. 19, pp. 387-393, 2005.
- [5] D. Arditi and T. Pulket, "Predicting the Outcome of Construction Litigation Using an Integrated Artificial Intelligence Model," *Journal of Computing in Civil Engineering*, vol. 24, pp. 73-80, 2010.
- [6] D. Martin Katz, et al., "A General Approach for Predicting the Behavior of the Supreme Court of the United States," *PLoS ONE*, vol. 12, pp. 1-18, 2017.
- [7] H. A. M. Shaffril, et al., "Systematic Literature Review on Adaptation towards Climate Change Impacts among Indigenous People in the Asia Pacific Regions," *Journal of Cleaner Production*, vol. 258, 2020.
- [8] H. Alsolai and M. Roper, "A systematic literature review of machine learning techniques for software maintainability prediction," *Inf. Softw. Technol.*, vol. 119, no. September 2019, p. 106214, 2020, doi: 10.1016/j.infsof.2019.106214.
- [9] H. Zhong, et al., "Legal Judgment Prediction via Topological Learning," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, pp. 3540-3549, 2020.
- [10] J. Li, et al., "Research and Design on Cognitive Computing Framework for Predicting Judicial Decisions," *Journal of Signal Processing Systems*, vol. 91, pp. 1159-1167, 2019.
- [11] K. J. Danjuma, "Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients," *IJCSI International Journal of Computer Science Issues*, vol. 12, 2015.
- [12] K. Nikolskaia and V. Naumov, "Artificial Intelligence in Law," *2020 International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon*, 2020.
- [13] K. W. Chau, "Prediction of Construction Litigation Outcome - A Case-based Reasoning Approach," *Lecture Notes*

- in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4031, pp. 548-553, 2006.
- [14] M. García-Jiménez, *et al.*, "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, pp. 1493-1516, 2020.
- [15] 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," *Research Synthesis Methods*, vol. 11, pp. 181-217, 2020.
- [16] M. Medvedeva, *et al.*, "Using Machine Learning to Predict Decisions of the European Court of Human Rights," *Artificial Intelligence and Law*, vol. 28, pp. 237-266, 2020.
- [17] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*. 2006.
- [18] N. Aletras, *et al.*, "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective," *PeerJ Computer Science*, vol. 2016, pp. 1-19, 2016.
- [19] N. Bagherian-Marandi, *et al.*, "Two-layered Fuzzy Logic-Based Model for Predicting Court Decisions in Construction Contract Disputes," *Artificial Intelligence and Law*, 2021.
- [20] P. Bhilare, N. Parab, N. Soni, B. Thakur, and A. Professor, "Predicting Outcome of Judicial Cases and Analysis using Machine Learning," *Int. Res. J. Eng. Technol.*, vol. 326, 2019.
- [21] R. Mothukuri and B. B. Rao, "Data Mining on Prediction of Crime and Legal Judgements: A State of an Art," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, pp. 3670-3679, 2019.
- [22] R. Sil and A. Roy, "A Novel Approach on Argument based Legal Prediction Model using Machine Learning," in *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020, Icosec*, pp. 487-490, 2020.
- [23] T. Mahfouz and A. Kandil, "Litigation Outcome Prediction of Differing Site Condition Disputes through Machine Learning Models," *Journal of Computing in Civil Engineering*, vol. 26, pp. 298-308, 2012.
- [24] V. Ben-David, "Study of Termination of Parental Rights: An Analysis of Israeli Court Decisions in Favor or Against Termination of Parental Rights," *Journal of Family Social Work*, vol. 18, pp. 225-242, 2015.
- [25] Z. Liu and H. Chen, "2017 SSCI proceedings," in *A predictive performance comparison of machine learning models for judicial cases*, 2017, pp. 1-6, doi: 10.1109/SSCI.2017.8285436.
-

A Systematic Literature Review of Machine Learning Methods in Predicting Court Decisions

ORIGINALITY REPORT

18%

SIMILARITY INDEX

PRIMARY SOURCES

- 1** Hayrol Azril Mohamed Shaffril, Nobaya Ahmad, Samsul Farid Samsuddin, Asnarulkhadi Abu Samah, Mas Ernawati Hamdan. "Systematic literature review on adaptation towards climate change impacts among indigenous people in the Asia Pacific regions", Journal of Cleaner Production, 2020
95 words — 2%
Crossref
- 2** [ascelibrary.org](https://www.ascelibrary.org)
Internet
94 words — 2%
- 3** "Big Data and Security", Springer Science and Business Media LLC, 2020
49 words — 1%
Crossref
- 4** Rafe Athar Shaikh, Tirath Prasad Sahu, Veena Anand. "Predicting Outcomes of Legal Cases based on Legal Factors using Classifiers", Procedia Computer Science, 2020
49 words — 1%
Crossref
- 5** [peerj.com](https://www.peerj.com)
Internet
42 words — 1%
- 6** www.tandfonline.com
Internet
42 words — 1%
- 7** www.mdpi.com
Internet

31 words — 1%

8 journals.vgtu.lt
Internet

30 words — 1%

9 Yueguo Zhang, Lili Dong, Xinghao Jiang. "Legal judgment prediction based on computer topology learning and fact elements", The 2nd International Conference on Computing and Data Science, 2021
Crossref

26 words — 1%

10 ZHE MIN XU, CHUN SUM NG. "ERROR PROBABILITY OF SPACE-TIME BLOCK CODED SYSTEMS IN MULTIPATH RAYLEIGH CHANNEL: MLSE AND SUBOPTIMUM DECODING", International Journal on Wireless & Optical Communications, 2007
Crossref

26 words — 1%

11 aclanthology.org
Internet

26 words — 1%

12 researchr.org
Internet

24 words — < 1%

13 journals.sagepub.com
Internet

23 words — < 1%

14 link.springer.com
Internet

23 words — < 1%

15 Riya Sil, Abhishek Roy. "A Novel Approach on Argument based Legal Prediction Model using Machine Learning", 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020
Crossref

22 words — < 1%

16	"Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020", Springer Science and Business Media LLC, 2021 Crossref	21 words — < 1%
17	idus.us.es Internet	20 words — < 1%
18	arxiv.org Internet	19 words — < 1%
19	www.vlkanova.sk Internet	19 words — < 1%
20	discovery.ucl.ac.uk Internet	17 words — < 1%
21	www.scilit.net Internet	17 words — < 1%
22	"Recent Advances in Intelligent Systems and Smart Applications", Springer Science and Business Media LLC, 2021 Crossref	16 words — < 1%
23	www.ncbi.nlm.nih.gov Internet	15 words — < 1%
24	María García-Jiménez, María Jesús Cala-Carrillo, M. Eva Trigo. "Predicting Disengagement from Judicial Proceedings by Female Victims of Intimate Partner Violence in Spain: A Systematic Replication With Prospective Data", Violence Against Women, 2019 Crossref	14 words — < 1%
25	eprints.utm.my Internet	14 words — < 1%

26 Yongjun Wang, Jing Gao, Junjie Chen. "Deep Learning Algorithm for Judicial Judgment Prediction Based on BERT", 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020 11 words — < 1%

Crossref

27 Hayrol Azril Mohamed Shaffril, Asnarulkhadi Abu Samah, Samsul Farid Samsuddin. "Guidelines for developing a systematic literature review for studies related to climate change adaptation", Environmental Science and Pollution Research, 2021 10 words — < 1%

Crossref

28 "Chinese Computational Linguistics", Springer Science and Business Media LLC, 2019 9 words — < 1%

Crossref

29 Choon Sen Seah, Shahreen Kasim, Mohd Farhan Md. Fudzee, Rohayanti Hassan. "Significant directed walk framework to increase the accuracy of cancer classification using gene expression data", Journal of Ambient Intelligence and Humanized Computing, 2020 9 words — < 1%

Crossref

30 www.iccep2019.com 9 words — < 1%

Internet

31 "Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing", Springer Science and Business Media LLC, 2021 8 words — < 1%

Crossref

32 Hongsong Dong, Fengbao Yang, Xiaoxia Wang, Yufeng Sun. "Automatic extraction of associated fact elements from civil cases based on a deep contextualized embeddings approach: KGCEE", Soft Computing, 2021 8 words — < 1%

33 Lufeng Yuan, Jun Wang, Shifeng Fan, Yingying Bian, Binming Yang, Yueyue Wang, Xiaobin Wang. "Automatic Legal Judgment Prediction via Large Amounts of Criminal Cases", 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019
8 words — < 1%
Crossref

34 Mahfouz, Tarek, Sukhrob Davlyatov, and Amr Kandil. "Analysis of Differing Site Condition (DSC) Litigation Reasoning Through Statistical Modeling", International Journal of Construction Education and Research, 2016.
8 words — < 1%
Crossref

35 Milton Campoverde-Molina, Sergio Lujan-Mora, Llorenc Valverde Garcia. "Empirical studies on web accessibility of educational websites: a systematic literature review", IEEE Access, 2020
8 words — < 1%
Crossref

36 Nidal Drissi, Sofia Ouhbi, Mohammed Abdou Janati Idtissi, Mounir Ghogho. "Gamification-Based Apps for PTSD: An Analysis of Functionality and Characteristics", 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), 2019
8 words — < 1%
Crossref

37 S. Jafar Ali Ibrahim, M. Thangamani. "Enhanced singular value decomposition for prediction of drugs and diseases with Hepatocellular carcinoma based on Multi-Source Bat Algorithm based random walk", Measurement, 2019
8 words — < 1%
Crossref

38 thesai.org
Internet
8 words — < 1%

39 "Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series", Springer Science and Business Media LLC, 2019

Crossref

7 words — < 1%

40 "Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2020

Crossref

7 words — < 1%

41 Sinasi Bektas, M. Talat Birgonul, Irem Dikmen. "Integrated Probabilistic Delay Analysis Method to Estimate Expected Outcome of Construction Delay Disputes", Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2021

Crossref

7 words — < 1%

42 etd.lib.metu.edu.tr

Internet

7 words — < 1%

43 Christopher Collins, Denis Dennehy, Kieran Conboy, Patrick Mikalef. "Artificial intelligence in information systems research: A systematic literature review and research agenda", International Journal of Information Management, 2021

Crossref

6 words — < 1%

44 Swasti Singhal, Laxmi Ahuja, Himanshu Monga. "State of The Art of Machine Learning for Product Sustainability", 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020

Crossref

6 words — < 1%

EXCLUDE QUOTES OFF

EXCLUDE MATCHES OFF

EXCLUDE BIBLIOGRAPHY ON