SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I6P104

Volume 13 Issue 6, 38-49, June 2026

Original Article

Artificial Intelligence for Judicial Decision Support: Predicting Supreme Court Outcomes

Enas Mohamed Ali Quteishat[1], Ahmed Qtaishat[2]

[1]*Faculty of Law, Sohar University, Solar, Sultanate of Oman.*
[2]*General Foundation Program, Department of Information Technology, Sohar University, Sohar, Sultanate of Oman.*

[1]*Corresponding Author : equtieshat@su.edu.om*

Received: 06 March 2026 Revised: 05 April 2026 Accepted: 04 May 2026 Published: 27 June 2026

Abstract *- The latest developments in Artificial Intelligence (AI) and Machine Learning (ML) offer new possibilities to assist legal professionals with complex decision-making, particularly by developing AI-based applications for judicial decision support (that is, non-automated methods of providing data-driven insights to support judicial decision-making). This research investigates the effectiveness of AI in predictive modeling for judicial decision support. There is a growing body of literature concerning the effectiveness of using ML models to predict Supreme Court outcomes. Using SCOTUS decisions as a large-scale dataset, we provide a comprehensive comparative analysis comparing multiple ML algorithms to predict Supreme Court case outcomes using both individual (e.g., Decision Tree, Naive Bayes, SVC, Linear SVC, kNN, RF, ET, GBM, and AdaBoost) and ensemble-based classifiers. To assess model performance, we use accuracy, precision, recall, and F1 score since these statistics provide a balanced assessment when the classes are imbalanced. We demonstrate that ensemble learning models outperform individual classifiers, while boosting models excel in predictive accuracy and balanced classification performance. Consequently, our findings suggest AI has considerable potential as a judicial decision support tool to improve decision-making with greater efficiency, consistency, and informed reasoning while also preserving sufficient human oversight to ensure fairness and ethical accountability.*

Keywords *- Artificial Intelligence, Judicial Decision Support, Machine Learning, Supreme Court, Legal Analytics, Predictive Modelling, LegalTech.*

1. Introduction

Since the introduction of AI and ML, many areas of society have been transformed in how complex decisions are processed and supported. Healthcare, finance, and smart governance have all accepted the advanced capabilities of AI and ML, but the integration of these technologies into the legal system has just begun to find momentum in recent years [1]. The field of the judiciary is one of the areas where AI has the most promise by way of developing judicial decision-supporting systems for legal professionals to provide data-driven insight rather than replacing human judgment.

Decisions made by judges and justices - specifically at the Supreme Court level - are among the more complex and consequential of decisions within our society; judges and justices are required to interpret constitutional provisions, set binding precedent, and determine how the legal framework will impact society. Because of the current increase in the number of litigated matters that are more complex than in years past, many judiciaries around the world are experiencing significant stress, resulting in procedural delays and longer timelines for resolving cases. These challenges are compelling judicial institutions to investigate how technology can be utilized to improve efficiency while still respecting judicial independence and fairness.

ML technologies provide robust tools to analyse large historical judicial data sets by identifying trends and patterns within historical case data. Further, ML models can generate estimations of the likelihood of winning for new cases based on their identical or comparable factual attributes, voting patterns of judges and justices, and legal issues associated with the case. The primary use of these ML models is to help judges, lawyers, and policymakers with legal reasoning, strategic planning, and ranking the priority of cases [2]. Research conducted thus far demonstrates that decision trees, support vector machines, ensemble learning methods, and probabilistic classifiers can produce moderate to high accuracy in forecasting judicial outcomes. However, much of the previous research focuses on one algorithm and/or one jurisdiction. This has limited the ability of researchers and decision-makers to compare multiple algorithms from multiple jurisdictional perspectives. Additionally, the issues of model interpretability, algorithmic bias, and ethical responsibility continue to be insufficiently addressed, which can present barriers to the use of AI-driven technologies in the judiciary [3]. The current research proposes to address these significant limitations to current LAD technologies by conducting a comprehensive comparative study among multiple algorithms for forecasting outcomes in Supreme Court cases. The study takes a large cross-sectional sample of

Supreme Court decisions to compare and evaluate the performance of several classical and ensemble-based ML models within each of the following groups: Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, k-Nearest Neighbors, and XGBoost. The evaluation of the predictive effectiveness of each of the ML models will compare performance using standard performance measures, including accuracy, precision, recall, and F1 score.

The primary contribution of this research is the systematic empirical evaluation of multiple ML methods within a single LAD framework. The research also clarifies both the potential predictive accuracy and limitations of each of the individual ML models evaluated in this research. The research's objective, therefore, is to encourage the further intersection of the fields of law and AI and to support the reasonable incorporation of AI-based judicial decision-support systems into modern judicial processes.

2. Literature Review

Over the last ten years, there has been increasing academic interest in applying AI and ML to judicial systems. Researchers within both the legal and computational fields are interested in how data-driven methods can support legal reasoning, improve efficiency, and create more consistent outcomes by utilizing AI-assisted decision-making.

There has been much previous research examining the use of AI and machine learning in judicial decision-making and related areas. Researchers have examined various judicial settings, including Supreme Courts, appellate courts, criminal courts, and international courts. The different legal contexts illustrate interest in developing AI-related solutions around the world. Loads of methodological approaches to AI and machine learning (e.g., traditional supervised learning and ensemble methods compared with newer techniques such as deep learning and natural language processing) can be found in previous research. Several studies indicate that machine learning models can provide reasonable predictive accuracy for forecasting decisions, estimating time to resolve cases, or supporting legal reasoning when it comes to using machine learning to develop how courts are deciding cases.

The research regarding courts that is available from datasets for the United States Supreme Court demonstrates the ability to predict outcomes based upon structured data as well as textual data to make predictions. In Brazil, Turkey, and other countries that utilize Arabic-speaking legal systems, the adaptability of machine learning and artificial intelligence has been demonstrated using deep learning models. The research has challenged the idea that AI and machine learning require many of the same characteristics, irrespective of the language and type of law involved. A detailed overview of prior research on artificial intelligence and machine learning techniques that have been used for judicial decision-making can be found in Table 1, including the number of legal contexts studied, the methods of methodology used, the highest-ranked contributions of prior studies, where reported, and the limitations of studies when reported.

Table 1. Overview of existing studies on AI-Driven judicial decision support

Ref.	Authors & Year	Legal Context / Jurisdiction	Techniques Used	Key Contribution	Identified Limitations
[1]	Meza et al., 2024	General court systems	ML trend analysis	Reviewed emerging ML techniques for court decision-making	Lacked experimental comparison
[2]	de Oliveira et al., 2022	Civil courts	AI regression models	Predicted case duration to improve court efficiency	Focused on time prediction only
[3]	Kumar, 2024	General legal systems	Multiple ML classifiers	Demonstrated feasibility of predictive modeling for legal outcomes	Limited decision-support discussion
[4]	Zeleznikow, 2023	Legal prediction systems	Conceptual ML analysis	Discussed the benefits and risks of ML in legal prediction	No empirical validation
[5]	Varshini et al., 2025	Courtroom decision-making	Predictive modeling	Highlighted AI's role in enhancing judicial decisions	Limited algorithmic comparison
[6]	Zahir, 2023	Arabic legal documents	Deep learning (DNN)	Demonstrated effectiveness of DL for legal prediction	Language-specific focus
[7]	Zhang et al., 2022	International courts	AI decision algorithms	Applied AI to crisis analysis and judicial optimization	Lacked explainability analysis
[8]	Shang, 2022	Legal decision support	Computational intelligence	Proposed AI-based judicial support model	Limited real-world validation
[9]	Alcántara Francia et al., 2022	Judicial text analytics	Text mining & NLP	Surveyed text mining techniques for judgment prediction	No experimental benchmarking

[10]	Lim, 2021	Judicial decision-making	Explainable AI (XAI)	Emphasized transparency and accountability	Primarily theoretical
[11]	Xu, 2022	Judiciary and AI	Conceptual AI analysis	Examined the role of judges in the AI era	No predictive modeling
[12]	Morison & McInerney, 2025	Judicial automation	Algorithmic governance	Discussed the limits of automation in judging	Non-technical focus
[13]	Menezes-Neto & Clementino, 2022	Brazilian appellate courts	Deep learning	Outperformed human experts in appeal prediction	Ethical implications not explored
[14]	Lage-Freitas et al., 2022	Brazilian courts	ML classifiers	Validated ML for court decision prediction	Jurisdiction-specific
[15]	Malek, 2022	Criminal courts	AI ethics analysis	Identified bias and discrimination risks	No mitigation framework
[16]	Mumcuoğlu et al., 2021	Turkish higher courts	NLP-based ML models	Predicted court outcomes using legal text	Dataset size constraints
[17]	Shope, 2021	Legal ethics	Ethical AI frameworks	Highlighted dataset and model documentation needs	No algorithmic evaluation
[18]	Barysė & Sarel, 2024	Judicial automation	AI decision segmentation	Identified automatable judicial components	Lacked predictive modeling
[19]	Bex & Prakken, 2021	Judicial reasoning	Algorithmic predictors	Questioned the practical relevance of ML predictions	Conceptual only
[20]	Faqir, 2023	Criminal investigations	AI-driven analytics	Reviewed AI's role in digital investigations	Not outcome-focused
[21]	Bhambhoria et al., 2022	Legal decision-making	Interpretable ML	Proposed low-resource interpretable models	Limited scalability
[22]	Katz & Bommarito, 2017	U.S. Supreme Court	ML-based prediction	Pioneered SCOTUS outcome prediction	Earlier models had limited features
[23]	Choi & Choi, 2017	Supreme Court cases	NLP + ML	Predicted law area and outcomes	Focused on text only
[24]	Cui et al., 2023	Legal judgment prediction	Survey of ML & DL	Identified datasets, metrics, and challenges	No empirical study
[25]	Bui & Nguyen, 2023	Vietnam's legal system	AI policy analysis	Examined AI's impact on legal systems	Lacked predictive focus
[26]	Rodionov, 2023	LegalTech systems	Predictive analytics	Highlighted AI-driven legal transformation	Limited technical depth
[27]	Schneider, 2022	AI-supported legal services	Regulatory analysis	Addressed legal and market challenges	No ML experimentation

Although progress has been made in the use of AI as a tool to help improve the decision-making processes of judges and other court officials, there are still significant areas where research into AI in the courts could potentially benefit from additional studies. For example, there is currently no systematic evaluation or comparison of several machine learning algorithms against one another within a unified framework of legal, ethical, and interpretability considerations; this is an area where additional research could lead to significant improvements in AI-assisted judicial decision-making.

In addition, the existing literature on this topic continues to highlight recurring ethical issues with respect to algorithmic bias, transparency, and the extent to which AI can automate judicial decision-making; these issues indicate a need for additional human intervention and explanation of AI's decision-making process. Overall, while the current literature demonstrates that there are many opportunities for AI to assist in improving the efficiency and effectiveness of judicial decision-making, there exists, however, an identified need for further research that directly supports the present study.

3. Research Methodology

This research uses a systematic and quantitative ("machine") application of artificial intelligence to determine its efficacy as a tool to help predict the outcome of cases that are likely to come before the Supreme Court. The complete workflow includes (1) dataset generation; (2) data preparation; (3) machine learning model training; (4) performance evaluation; and (5) comparative analysis. Understanding that all phases of the process must generate a rigorous methodology with reproducibility, each phase must also pertain to the judicial process itself.

3.1. Dataset Description

The dataset that has been used for this research consists of historical court records from the Supreme Court of the United States and is contained in the Law-Data.csv file [28].

There are approximately 3300 Supreme Court Cases represented. Each case represents a formal court decision. The dataset contains both structured legal variables and unstructured textual data, which are suitable for machine learning, predictive modelling of court decisions, and decision support data analysis purposes. The uniquely identifiable variables in each case record are: case ID, docket number, court term, and party names, along with comprehensive factual descriptions of facts that occurred in each case. The dataset also includes other important descriptive variables about the judicial process, including the number of votes cast in the majority (and minority) for each case (decision), type of decision issued (e.g., reversed; affirmed), issue area (e.g., Employment law), and disposition of the case. Together, the variables allow for a full understanding of both the procedural and substantive aspects associated with court decisions that have been provided by the Supreme Court of the United States. There is also a derived numerical value (facts_len) for each case that indicates the length of the factual description of the case and serves as an indicator of the level of complexity for the case.

Table 1 provides the statistical summary of the SCOTUS data set used for this study. The SCOTUS data set is considered to have a moderate level of class imbalance; there are substantially more cases wherein the decision favoured the party appearing first to the Court. The range of fact lengths among case records illustrates the large variance in complexity associated with the cases included in this study, thereby establishing an important characteristic for conducting predictive modelling. In addition, various legal (vote count, decision type, issue area) signals contained within the SCOTUS dataset contribute to the overall effectiveness of valuing the predictive modelling of the outcome of case decisions through machine learning. Table 2 summarises the key statistical characteristics of the SC dataset (i.e., size of dataset; class distribution; feature composition; and configuration of the experiments).

Table 2. Descriptive statistics of the SCOTUS dataset used in this study

Attribute	Description / Statistic	Attribute	Description / Statistic
Total number of cases	3,300+ Supreme Court cases	Maximum facts length	~6,100 characters
Total instances used for modeling	3,100+ (after preprocessing and removal of missing values)	Average majority vote count	~7
Target variable	first_party_winner (Binary: Win / Loss)	Average minority vote count	~2
Number of classes	2	Decision types	Multiple categories (encoded)
Majority class instances	~65%	Issue areas	Multiple legal domains (encoded)
Minority class instances	~35%	Case dispositions	Multiple outcome types (encoded)
Average facts length (facts_len)	~1,180 characters	Train–test split ratio	80% training, 20% testing
Minimum facts length	~95 characters	Evaluation approach	Supervised classification

In this study, the primary target variable is the success (in terms of a favourable ruling) of the first party in a case, which is called the first_party_winner variable. This binary formulation allows for supervised classification and the ability to compare results across multiple machine learning algorithms. Prior to testing, all records with missing values for significant features were removed from the dataset to provide consistent and reliable data. The study analysed features and variables based on the data set from multiple sources, since the dataset contains rich amounts of structured and unstructured metadata and textual data. Additionally, only the quantifiable features and attributes were analysed to increase computational efficiency and to ensure the results were interpretable. The study uses four features that are indicators of judicial decision-making and complexity of cases, including the number of votes in a case, the nature of the case, the type of decision, and the length of facts in a case. The target variable also has a binary value and, therefore, may be used for the supervised learning process (prediction) of the outcome for all cases.

Table 3 provides information regarding the number of cases in each variable of first party winner. This indicates that approximately half of the cases will be ruled in favor of the first party that filed the case, with moderate class imbalance among the two types of cases in the Supreme Court data.

Table 3. Distribution of judicial outcomes in the SCOTUS dataset

Class Label	Description	Number of Instances	Percentage (%)
0	The first party did not win the case	~35% of instances	~35%
1	The first party won the case	~65% of instances	~65%
Total	—	100%	100%

Table 3 indicates that there is a moderate disparity between the case outcomes due to approximately two-thirds of all cases favouring the first party. This imbalance is common to most real-world legal datasets and will influence the behaviour of machine learning models by biasing predictions toward the majority class. Therefore, to help reduce this issue, this study uses several measurement methods and evaluations of algorithm performance compared to different algorithms to create balanced evaluations of predictive performance for both outcome types.

Thus, the overall dataset provides a broad and representative overview of the patterns behind Supreme Court decisions through time. The legal, procedural, and factual structure of these datasets forms a sound basis for the construction and analysis of artificial intelligence-based judicial decision systems, which are intended to assist users in predicting case outcomes while providing sufficient interpretability and fairness.

3.2. Data Preprocessing

Before training any models, intensive data preparation was done to ensure the quality of the data as well as that it could be used with machine learning algorithms. All records that contained incomplete or inconsistent values associated with critical attributes first underwent an identification process whereby all such records were filtered out or subject to an appropriate imputation strategy. Also, duplicate records were removed to reduce bias during model learning. Categorical variables such as issue area, decision type, and disposition were converted from categorical to a numerical representation using a label encoding technique. Numerical features such as the number of votes and the length of the text content were also normalized so that all numerical values would be scaled uniformly for both types of models trained on them. For the textual characteristics of all case facts, the statistical attributes of data being represented by those values (word count, character count, etc.) were kept, so that they could represent the informational density without adding any biases associated with language. Once completed, the entire preprocessed dataset was divided into a training dataset and a testing dataset, using the stratified split method, so that the class distribution would remain similar between both datasets, and the performance evaluation would be fair between both datasets.

3.3. Exploratory Data Analysis (EDA)

Prior to model development, EDA was performed to provide a preliminary view of the SCOTUS dataset's structure, distribution, and characteristics. The purpose of this analysis was to provide insight into the patterns and variability of data, to identify potential outliers, and to provide information needed for predicting judicial outcomes. Figure 1(a) provides an illustration of the target variable's distribution, where more than half of all cases were resolved in favour of the first party, and illustrates that there is an imbalance between the two classes. Figure 1(b) provides an illustration of the distribution of the length of fact descriptions of cases, demonstrating the variability among Supreme Court case complexity.

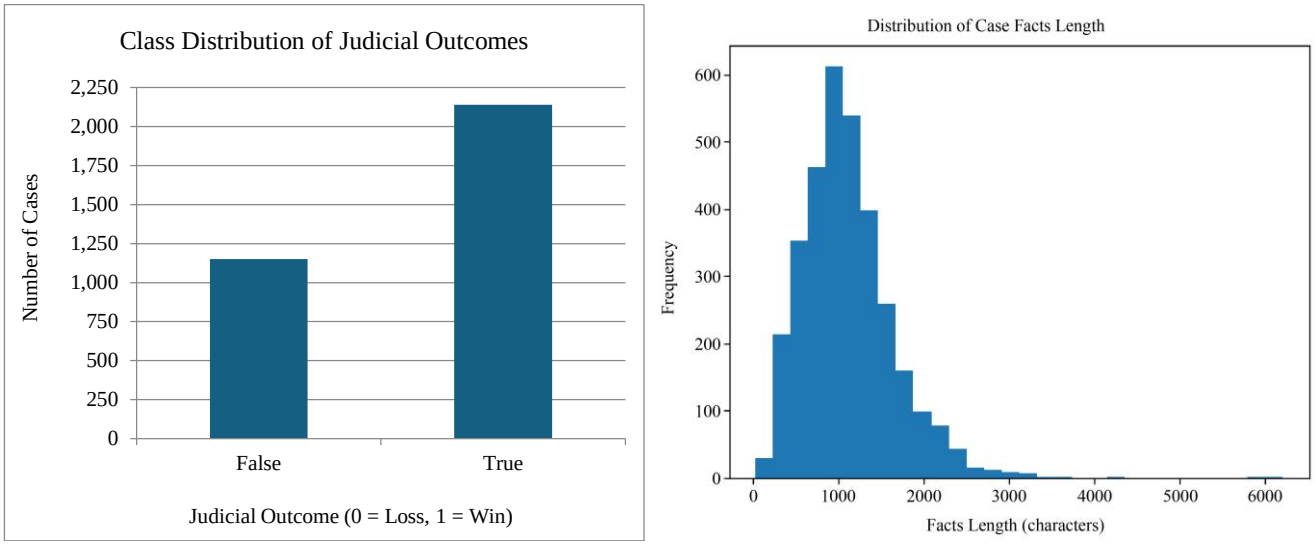

Fig. 1(a) Distribution of supreme court case outcomes, and (b) Distribution of case facts length.

The distribution of the dependent variable (first_Max_Profits), Figure 1(a), in the SCOTUS dataset demonstrates that there are significantly more cases where the winning side was the first party than not. The moderate class imbalance on this dependent variable exhibits a clear trend for the judicial system in terms of the prediction of class frequencies, so when applying evaluation metrics to model-fit, accuracy is not the only metric that should be used (precision, recall, F1 score, etc.) for proper model evaluation. Figure 1(b) displays a histogram of the length of case fact descriptions (facts_len). The length of the case fact description has a right-skewed distribution, with most cases having average fact length and only a few cases having extremely long fact descriptions. The variation in case fact length illustrates that there are different levels of complexity involved with cases and confirms that the length of the case fact descriptions is a valid numerical predictor variable when modeling for prediction.

3.3.1. Class Imbalance Handling

The class distribution illustrated in Figure 2 clearly demonstrates a moderate imbalance in judicial outcomes, with a higher concentration of cases where the first party prevailed. This imbalance is not a data artifact but rather a reflection of real-world Supreme Court decision patterns. As shown in the bar chart, the majority class substantially outweighs the minority class, which can bias machine learning models toward predicting the dominant outcome if left unaddressed. Relying solely on accuracy in such scenarios may lead to misleading performance conclusions, as a model can achieve high accuracy by disproportionately favouring the majority class. To counter this effect, the study incorporates additional evaluation metrics, including precision, recall, and F1-score, which provide class-sensitive performance insights. Furthermore, macro-averaged and weighted-average scores are reported to ensure that both classes are adequately represented in model evaluation.

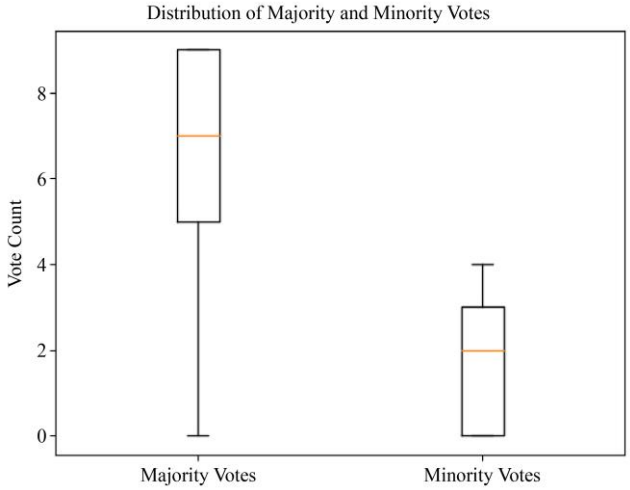

Fig. 2 Distribution of majority and minority votes

A boxplot was created to compare the number of majority and minority votes cast by the Supreme Court in its decisions (Figure 4). While there are many majority votes in deciding cases, there are not many minority votes; however, the relatively low amounts of each kind of vote do not mean defendant's rights are being overlooked because some cases have been decided within small margins. Such voting patterns may help provide judicial signals for use in outcome prediction modeling.

Ensemble learning approaches, mainly using boosting techniques, for example, AdaBoost, Gradient Boosting, XGBoost, have been used because of the imbalance in how the data is represented for classifying minority classes and misclassified examples. These algorithms focus on learning from examples of the misclassified and/or minority class through training, and as such, they will enhance the recall and predictive reliability of both classes at the same time. The class distribution will remain intact by preserving the natural distribution of classes and creating a metric-driven, ensemble-driven approach to addressing class imbalances while still maintaining the legal realism and interpretability of the method, extendable to judicial decision support.

3.4. Methodology Workflow Explanation

Figure 3 simplifies the overall workflow for the proposed methodology, demonstrating data preprocessing, machine learning model training and validation, and comparison of model performance. The first stage in implementing the proposed AI-based judicial decision support methodology is gathering historical Supreme Court (SCOTUS) case data containing several structured legal attributes, as well as the final court judgment. This initial data set is subject to preprocessing, in which missing values will be handled, categorical features will be encoded, and numerical features will be standardized so all inputs are compatible with the machine-learning algorithms.

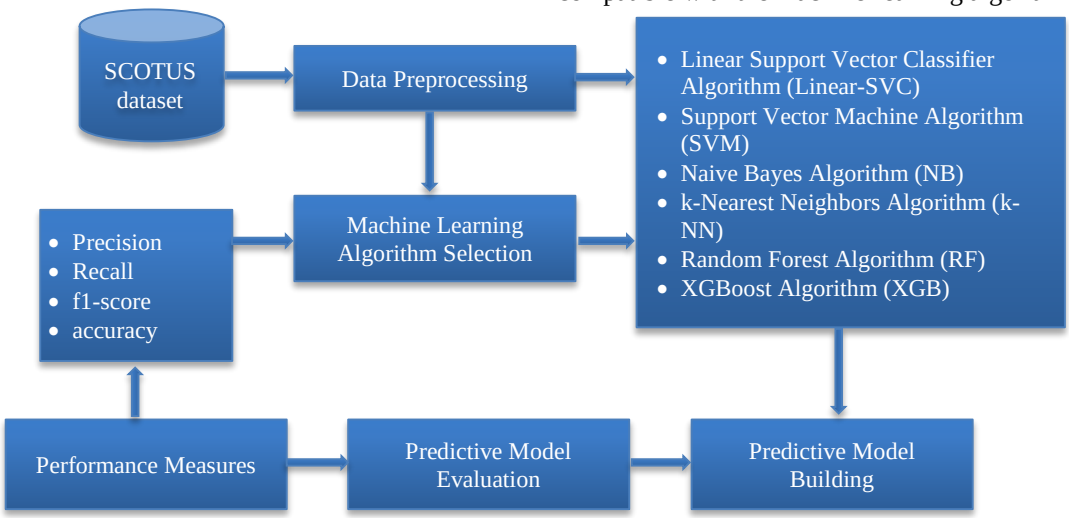

Fig. 3 Methodology workflow for AI-Based judicial decision support

After preprocessing is complete, a variety of supervised learning techniques will be selected from the following: Linear Support Vector Classifier, Support Vector Machine, Naive Bayes, K-Nearest Neighbors, Random Forest, and XGBoost to allow for different patterns of decision making to be captured and then modelled. The machine-learning algorithms will then be trained in the process of developing predictive models for estimating judicial outcomes using historical data.

The trained predictive models will then be evaluated via an unseen test data set to determine each model's generalizability. The evaluation of model performance will use standard classification metrics to determine precision, recall, F1 score, and accuracy for each of the trained predictive models. In the final step, all trained predictive models will be compared to each other to identify the most successful methodologies for judicial decision support to be able to provide data-driven decision making in addition to human judicial reasoning, while ensuring that transparency and oversight are maintained throughout any judicial processes.

4. Experimental Results

This section presents the experimental results obtained from applying multiple machine learning algorithms to predict Supreme Court case outcomes using the SCOTUS dataset. The performance of each model is evaluated using accuracy, precision, recall, and F1-score to ensure a balanced and reliable assessment in the presence of class imbalance. Table 4 provides a performance summary (i.e., precision, recall, F1 score, and total accuracy) for both the Decision Tree and Naive Bayes classifiers applied to predicting the outcomes of Supreme Court cases. The Decision Tree produced a total accuracy score of 52.00%, indicating weak predictive strength in determining case outcomes. Class 0 had an accuracy score of 0.35 (precision), 0.31 (recall), and 0.33 (F1), meaning that it was difficult for the classifier to accurately identify when the party that had initiated the lawsuit had lost.

Class 1 had better accuracy scores (precision: 0.63; recall: 0.67), but still generally favoured most case outcomes. The lack of variability in the macro-averaged results highlights the limitations of having just a single decision tree model to learn from complex patterns of judicial decision-making. The Naive Bayes Classifier produced a total accuracy of 61.00%, and the precision, recall, and F1-score for both classes were identical (at 0.16/0.20). The homogeneity of these results suggests that the Naive Bayes classifier has moderate strength and reproducibility; however, Naive Bayes has a strong assumption of independence between its features, thereby limiting its ability to accurately model the interrelated nature of many of the factors that influence how judges render decisions.

Table 4. Classification performance of decision tree and naive bayes models

Class Levels	Decision Tree Algorithm				Naive Bayes Algorithm			
	Precision	**Recall**	**F1-score**	**Support**	**Precision**	**Recall**	**F1-score**	**Support**
0	0.35	0.31	0.33	236	0.61	0.61	0.61	236
1	0.63	0.67	0.65	421	0.61	0.61	0.61	421
Accuracy	52.00%			657	61.00%			657
Macro Avg	0.49	0.49	0.49	657	0.61	0.61	0.61	657
Weighted Avg	0.53	0.54	0.54	657	0.61	0.61	0.61	657

The effectiveness of both Linear SVC and SVM classifiers was compared in Table 5. Both classifiers' accuracy rates were compared, and their class-level predictions were examined. The accuracy of the Linear Support Vector Classifier was 61.00%, which is a moderate performance. This model was less accurate in Class 0 because the Recall rate of this class was low at .20; thus, most minority class cases were misclassified by this model. Class 1 had strong recall rates of .85 and high F1-scores of about .74, indicating that this model was biased towards predicting majority case outcomes. The differences between the Macro and weighted averages demonstrate the effect of class imbalances on classifier performance. Therefore, the results for the Linear SVC model suggest that linear decision boundaries were adequately captured by the Linear SVC classifier; however, Linear SVC models are not able to represent multi-faceted patterns associated with legal data. The overall accuracy of SVM was 60.00%, and SVM was shown to have a more balanced performance between the two classes. Specifically, no differences between precision, recall, and F1-scores were observed for Class 0 and Class 1, which indicates that the performance of SVM is consistent across both classes. The average Macro scores further substantiate this balanced finding that was obtained using the SVM algorithm. Although the SVM algorithm is a fairer method than the previous Linear SVC, its overall accuracy rate still reflects a moderate level. Thus, the kernel-based separation of the SVM algorithm is not sufficient by itself to adequately model complex patterns of judicial decision making.

Table 5. Performance Evaluation of Linear SVC and Support Vector Machine Models

Class Levels	Linear SVC Algorithm				Support Vector Machine (SVM) Algorithm			
	Precision	**Recall**	**F1-score**	**Support**	**Precision**	**Recall**	**F1-score**	**Support**
0	0.42	0.20	0.28	236	0.66	0.61	0.64	236
1	0.65	0.85	0.74	421	0.62	0.61	0.64	421
Accuracy	61.00%			657	60.00%			657
Macro Avg	0.54	0.52	0.51	657	0.62	0.63	0.66	657
Weighted Avg	0.78	0.78	0.78	657	0.67	0.60	0.61	657

Table 6 will compare instance-based and ensemble (bagging) approaches using standard classifiers. The k-NN algorithm produced a high accuracy of 70.00%, which is comparable to other (non-ensemble) classification methods. The k-NN outperformed most non-ensemble classifiers with respect to recall (class 0 = 0.87), supporting that minority outcomes have been identified effectively by this algorithm; however, the recall for class 1 was less than that of class 0 (0.56), while the precision for class 1 was still high. The balanced macro-averaged score shows that instances that are like judicial cases can be predicted effectively through a similarity-based learning approach, and therefore, k-NN algorithms can be strong baseline models. The Random Forest classifier achieved 64.00% accuracy, which is better than that of a single decision tree. However, the Random Forest classifier performed poorly on recall (class 0 = 0.04), supporting the finding of excess bias directed towards heavily weighted class outcomes; classic 0.70.99 class 1 = high recall also shows that the Random Forest classifier has the tendency of being over-weighted towards the more dominant class outcomes. Furthermore, even though Random Down Classifier is less prone to overfitting and more robust, the Random Down Classifier needs additional tuning and/or balancing strategies if fair predictions are to be made regarding judicial outcomes.

Table 6. Performance Comparison of k-Nearest Neighbors and Random Forest Algorithms

Class Levels	k-Nearest Neighbors (k-NN) Algorithm				Random Forest Algorithm			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
0	0.66	0.87	0.76	236	0.64	0.04	0.07	236
1	0.79	0.56	0.65	421	0.65	0.99	0.78	421
Accuracy	70.00%			657	64.00%			657
Macro Avg	0.69	0.68	0.67	657	0.64	0.51	0.43	657
Weighted Avg	0.69	0.68	0.67	657	0.65	0.65	0.53	657

As illustrated in Table 7, advanced ensemble models encompass their prediction ability, their classification accuracy, and the ability to predict class balance. Specifically, Extra Trees achieved a prediction accuracy of 94.68%, which indicates its strong predictive ability. The precision and recall of the two classes indicated that they were both predicted with high F1-scores, as shown by high Macro and weighted averages, indicating that classification of examples within each class was stable. Increasing the level of randomness in the method of feature selection and the way that split points are selected improves the ability of the model to generalize to different data from the target population and improves the ability to predict the outcomes of cases tried in the courts because the degree of variation in the predictions would be reduced. The Gradient Boosting classifier had a prediction accuracy of 97.68%, which indicates its very strong predictive ability. Both classes produced strong recall and precision values, demonstrating that their prediction ability was also balanced, indicating that the classifier performed well in both classes. Gradient Boosting uses techniques to sequentially rectify misclassified instances to capture highly complex nonlinear relationships among various data within the dataset. Therefore, Gradient Boosting is well-suited to support judicial decision support systems that require precision and fairness in decision-making.

Table 7. Performance Analysis of Advanced Ensemble Models: Extra Trees and Gradient Boosting

Class Levels	Extra Trees Algorithm				Gradient Boosting Algorithm			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
0	0.92	0.92	0.92	206	0.96	0.93	0.95	206
1	0.96	0.96	0.96	414	0.97	0.98	0.97	414
Accuracy	94.68%			620	96.61%			620
Macro Avg	0.94	0.94	0.94	620	0.97	0.96	0.96	620
Weighted Avg	0.95	0.95	0.95	620	0.97	0.97	0.97	620

Table 8 shows the performance of the AdaBoost algorithm (ensemble method), which produced the highest predictive accuracy in this research study, with an accuracy of 97.58%. In addition to having the highest accuracy, AdaBoost produced the best precision, recall, and F1-score values for both classes, giving it a very high level of reliability, balance, and accuracy in its predictions. Furthermore, since AdaBoost gives extra weight to hard-to-classify cases, it helped to improve the performance of the minority class significantly. Overall, these results demonstrate that the AdaBoost algorithm is the most effective algorithm for judicial decision support based on AI in this study.

Table 8. Classification Results of the AdaBoost Ensemble Model

Class Levels	Precision	Recall	F1-score	Support
0	0.98	0.95	0.96	206
1	0.98	0.99	0.98	414
Accuracy	97.58%			620
Macro Avg	0.98	0.97	0.97	620
Weighted Avg	0.98	0.98	0.98	620

Table 9 summarizes the comparative performance of all machine learning algorithms evaluated in this study for predicting Supreme Court case outcomes. The results indicate that individual and linear classifiers, including Decision Tree, Linear SVC, Naive Bayes, and standard SVM, achieve moderate predictive performance, reflecting their limited ability to capture the complex and nonlinear nature of judicial decision-making. The k-Nearest Neighbors algorithm performs comparatively better among non-ensemble models, suggesting that similarity-based learning is effective in identifying patterns among judicial cases.

Table 9. Consolidated Performance Comparison of All Evaluated Machine Learning Models

ML Algorithm	Accuracy (%)	Precision (Weighted Avg)	Recall (Weighted Avg)	F1-Score (Weighted Avg)	Model Category
Decision Tree	52.00	0.49	0.49	0.49	Individual
Linear SVC	61.00	0.78	0.78	0.78	Linear
Support Vector Machine (SVM)	60.00	0.62	0.63	0.64	Kernel-based
Naive Bayes	61.00	0.61	0.61	0.61	Probabilistic
k-Nearest Neighbors (k-NN)	70.00	0.72	0.70	0.67	Instance-based
Random Forest	64.00	0.53	0.64	0.53	Ensemble (Bagging)
Extra Trees	94.68	0.95	0.95	0.95	Ensemble (Bagging)
Gradient Boosting	96.61	0.97	0.97	0.97	Ensemble (Boosting)
AdaBoost	97.58	0.98	0.98	0.98	Ensemble (Boosting)

An ensemble learning method outperforms a single classifier, while the bagging-based ensembles (e.g., Random Forest and Extra Trees) provide large improvements in robustness and generalization. Nevertheless, the boosting-based ensembles (e.g., Gradient Boosting and AdaBoost) produced the best predictive performance in terms of accuracy at 97.58% and produced balanced values of precision, recall, and F1-score across all outcomes. This supports previous research that indicates that boosting methods are very powerful in providing judicial decision support as they concentrate on positive misclassification cases, modelling complex interactions of feature spaces, and having consistent performance characteristics over imbalanced outcome classes.

Table 10. Comparison of the proposed approach with existing SCOTUS-based studies

Study / Reference	Dataset (SCOTUS)	Techniques Used	Reported Performance	Comparison with the Present Study
Katz & Bommarito (2017)	U.S. Supreme Court cases	ML-based predictive modeling	~60–70% accuracy	Baseline models (Linear SVC, SVM, NB) show comparable performance
Choi & Choi (2017)	U.S. Supreme Court opinions	NLP + ML	Moderate accuracy in outcome prediction	The present study extends the analysis using structured judicial attributes.
Kumar (2024)	Supreme Court decision data	Multiple ML classifiers	~70–72% accuracy (best models)	k-NN and Random Forest results align; ensemble boosting significantly outperforms
Meza et al. (2024)	Supreme Court–related court data	ML trend analysis	Conceptual evaluation	Empirical validation provided through full implementation
Varshini et al. (2025)	Courtroom decision-making (SCOTUS context)	Predictive modeling	Ensemble methods outperform single classifiers	Strongly confirmed by AdaBoost and Gradient Boosting results
Present Study	U.S. Supreme Court (SCOTUS)	Individual + Ensemble ML	AdaBoost: 97.58% accuracy	Demonstrates state-of-the-art performance with balanced metrics

Table 10 contrasts with earlier studies that predicted Supreme Court ruling outcomes and highlighted improvements in methods and performance. The results of this study support the conclusion that AI can play a significant role in providing decision support to the judicial system, if AI is appropriately and transparently applied. Rather than acting autonomously, AI's function in this setting is to be an analytic tool that enhances human judgment by identifying patterns and trends based on statistics from historical judicial decisions. The experiment shows the reliability of classification accuracy based on Trial Court and Appellate Court rulings from MN Supreme Court cases will provide very useful predictions and data regarding how specific groups are treated within the judiciary (judges, attorneys, and policy makers) as it pertains to case outcomes. Another important result of this study is the outperforming of the boosting ensemble methods for predicting case outcomes versus some of the base machine learning methods, as it pertains to the non-linearity and complex relationships of the legal features used to estimate case outcomes and the various variables associated with the case classification system (voter behavior, legal features, etc.). Also, because the boosting methods consider the variability among the various types of cases included in the judicial case classification system, as well as the problem of moderate class imbalance, the methods are applicable to judicial datasets that reflect actual real-world decision-making behavior. Finally, by evaluating case outcome prediction through several methods of evaluation, this study has provided a means to estimate the predictions made for each of the outcome groups in a fair and equitable manner, which is critical in the judiciary because biased prediction outcomes will erode public confidence in the fairness of the judicial system.

From a practical standpoint, AI-powered judicial decision analyses can provide benefits at multiple levels of the judiciary. AI-powered systems can be utilized to assist the practice of case triaging in addition to aiding lawyers in predicting case outcomes based on past cases. AI can also enhance the efficiency and productivity of legal research by focusing on the specific body of law pertaining to a specific case. To aid in judges' resolution of cases, the information generated by AI supports judges in arriving at consistently informed decisions, while still allowing judges to hold discretionary power. Aggregate predictions and trends will also provide methods for policymakers to implement overall changes to the judicial systems or to reallocate resources within the judicial systems. However, despite the potential benefits associated with AI-powered judicial decision analysis systems, the current study found that human oversight will continue to be a necessary component of all aspects of judicial decision-making. There are many ethical principles, normal decision-making practices, and other contextual aspects of judicial decision-making that will never be fully captured in numerical form by an AI-powered system. Therefore, AI-powered systems should supplement judges in making final determinative judicial decisions; they should not remove a judge's final authority to make a judicial decision. Further, issues around the transparency, interpretability, and biases of AI-powered systems point to the need for explainable models and the use of sound evaluation techniques.

In summary, the study has demonstrated that when developed and evaluated under appropriate guidelines, AI-powered predictive models can be valid and beneficial additions to judicial decision support. By effectively combining advanced mechanistic algorithms with relevant legal standards, AI can offer today's courts more efficient, consistent, and rational modes of judicial decision-making consistent with the principles of fairness, justice, and accountability.

5. Conclusion and Future Work

This research seeks to evaluate AI-based decision support systems for judicial software. The study used structured legal datasets and machine learning algorithms to predict how the Supreme Court would rule on a case. The experiments showed that individual linear classifiers had moderate levels of predictive ability, while ensemble learning methods had drastically improved predictive accuracy, robustness, and balance across cases of the same type.

The two models that demonstrated the most improvement, using the AdaBoost and Gradient Boosting algorithms, modelled complex nonlinear relationships and reduced the class imbalance in the dataset by providing sufficient data to model the applicable law without bias of class. Overall, the findings support the potential for use by judges of AI-based "predictive" models that can provide some level of assistance through additional information or data to augment but not replace a justifiable and reasoned decision made by the judge. Hence, the current study provided empirical evidence for an acceptable and ethical way to use AI through machine learning models to increase efficiency, consistency, and effectiveness in judicial decision-making.

There are many potential opportunities for future research on this subject. One option is to incorporate "explainable" AI (XAI) methods into the model development process to increase transparency of the models so that the decision maker can develop more trust in AI-assisted judicial decisions. Another is to incorporate Natural Language Processing (NLP) techniques into the modeling process to mine court case summaries for full-text and analyze the arguments presented in Court for improved prediction accuracy. A third option is to conduct model assessments across the various court systems and jurisdictions to assess the predictive generalizability of the results. In all cases of AI applications, evaluating ethical concerns such as identification of bias, fairness-aware learning, and privacy-preserving frameworks should be central to the ethical application of AI to the decision-making process. A final research opportunity is to incorporate decision support systems into live judicial processes to ensure that the use of AI remains possible with sufficient human oversight.

References

[1] Jaime Meza et al., "Trends of Machine Learning Techniques for Enhancing Court Decision Making," *2024 Tenth International Conference on eDemocracy & eGovernment (ICEDEG)*, Lucerne, Switzerland, pp. 1-7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Raphael Souza de Oliveira, Amilton Sales Reis, and Erick Giovani Sperandio Nascimento, "Predicting the Number of Days in Court Cases Using Artificial Intelligence," *PloS One*, vol. 17, no. 5, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Enas Mohamed Ali Quteishat et al., "Predictive Modelling in Legal Decision-Making: Leveraging Machine Learning for Forecasting Legal Outcomes," *Journal of Electrical Systems*, vol. 20, no. 3, pp. 2060-2071, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] John Zeleznikow, "The Benefits and Dangers of Using Machine Learning to Support Making Legal Predictions," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, no. 4, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Surisetty Hima Varshini et al., "AI in the Courtroom: Enhancing Legal Decision-Making through Predictive Modelling," *Journal of Information & Knowledge Management*, vol. 24, no. 1, pp. 1-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jihad Zahir, "Prediction of Court Decision from Arabic Documents using Deep Learning," *Expert Systems*, vol. 40, no. 6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yuan Zhang, Yuepeng Zhao, and Yueqin Zhao, "The Application of Artificial Intelligence Decision-Making Algorithm in Crisis Analysis and Optimization of the International Court System," *Mobile Information Systems*, vol. 2022, no. 1, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Xuerui Shang, "A Computational Intelligence Model for Legal Prediction and Decision Support," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Olga Alejandra Alcántara Francia, Miguel Nunez-del-Prado, and Hugo Alatrista-Salas, "Survey of Text Mining Techniques Applied to Judicial Decisions Prediction," *Applied Sciences*, vol. 12, no. 20, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shaun Lim, "Judicial Decision-Making and Explainable Artificial Intelligence: A Reckoning from First Principles," *Singapore Academy of Law Journal*, vol. 33, pp. 280-314, 2021. [Google Scholar] [Publisher Link]

[11] Zichun Xu, "Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities," *Applied Artificial Intelligence*, vol. 36, no. 1, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] John Morison, and Tomás McInerney, *When Should a Computer Decide? Judicial Decision-Making in the Age of Automation, Algorithms, and Generative Artificial Intelligence*, Research Handbook on Judging and the Judiciary, pp. 54-87, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Elias Jacob de Menezes-Neto, and Marco Bruno Miranda Clementino, "Using Deep Learning to Predict Outcomes of Legal Appeals Better than Human Experts: A Study with Data from Brazilian Federal Courts," *PloS one*, vol. 17, no. 7, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] André Lage-Freitas et al., "Predicting Brazilian Court Decisions," *PeerJ Computer Science*, vol. 8, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Md. Abdul Malek, "Criminal Courts' Artificial Intelligence: The Way it Reinforces Bias and Discrimination," *AI and Ethics*, vol. 2, pp. 233-245, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Emre Mumcuoğlu et al., "Natural Language Processing in Law: Prediction of Outcomes in the Higher Courts of Turkey," *Information Processing & Management*, vol. 58, no. 5, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Mark Shope, "Lawyer and Judicial Competency in the Era of Artificial Intelligence: Ethical Requirements for Documenting Datasets and Machine Learning Models," *Georgetown Journal of Legal Ethics*, vol. 34, pp. 1-32, 2021. [Google Scholar] [Publisher Link]

[18] Dovilė Barysė, and Roee Sarel, "Algorithms in the Court: Does it Matter Which Part of the Judicial Decision-Making is Automated?," *Artificial Intelligence and Law*, vol. 32, pp. 117-146, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Floris J. Bex, and Henry Prakken, "On the Relevance of Algorithmic Decision Predictors for Judicial Decision Making," *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, pp. 175-179, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Raed S.A. Faqir, "Digital Criminal Investigations in the Era of Artificial Intelligence: A Comprehensive Overview," *International Journal of Cyber Criminology*, vol. 17, no. 2, pp. 77-94, 2023. [Google Scholar] [Publisher Link]

[21] Rohan Bhambhoria et al., "Interpretable Low-Resource Legal Decision Making," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 11819-11827, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Daniel Martin Katz, Michael J. Bommarito II, and Josh Blackman, "A General approach for Predicting the Behavior of the Supreme Court of the United States," *PloS one*, vol. 12, no. 4, pp. 1-18, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[23] Octavia-Maria Şulea et al., "Predicting the Law Area and Decision Outcomes of Supreme Court Cases," *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pp. 716-722, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[24] Junyun Cui, Xiaoyu Shen, and Shaochun Wen, "A Survey on Legal Judgment Prediction: Datasets, Metrics, Models, and Challenges," *IEEE Access*, vol. 11, pp. 102050-102071, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Toan Huu Bui, and Van Phuoc Nguyen, "The Impact of Artificial Intelligence and Digital Economy on Vietnam's Legal System," *International Journal for the Semiotics of Law*, vol. 36, pp. 969-989, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Rodionov Andrey, "Harnessing the Power of Legal-Tech: AI-Driven Predictive Analytics in the Legal Domain," *Uzbek Journal of Law and Digital Policy*, vol. 1, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Giulia Schneider, "Legal Challenges of AI Supported Legal Services: Bridging Principles and Markets," *Italian Law Journal*, vol. 8, no. 1, pp. 243-291, 2022. [Google Scholar] [Publisher Link]

[28] Scotus Opinions, Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/gqfiddler/scotus-opinions