Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020)
IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

Verdict Prediction for Indian Courts Using Bag of Words and Convolutional Neural Network

V Gokul Pillai
gokulpillai05@gmail.com
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, Amritapuri, India

Lekshmi R Chandran
lekshmichandran@am.amrita.edu
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, Amritapuri, India

***Abstract*— All courts in India publish judgment by statistically analyzing the data of different cases and understanding the verdict from precedents judgments and the statute law. The approach of our study is to highlight the importance of Convolutional Neural Network (CNN) and Natural Language Processing (NLP) in the legal domain. The Bag of Words technique is being proposed i.e., one of the NLP tool to analyze the text of the court proceedings to extract the keywords from the text and CNN to classify each case into its charges (as per judicial law of India), to predict whether it is a bailable or a non-bailable offence and to give an approximate judicial decision. The results show that this method has an average accuracy of 85% in prediction based on the IPC (Indian Penal Code) which is extracted from the case files. This research work data is taken from the judicial pronouncement and the constitution of India.**

Keywords—Legal intelligence, Deep learning, Indian Law of Justices, Natural Language Processing, Bag of Words, Convolutional Neural Network.

I. INTRODUCTION

In the 21st century, technology is crawling into many facets of our lives. Imagine a situation where every judicial decision is taken by a robot instead of a human being. It seems daunting, doesn't it? But look at things from another perspective, it has a brighter side as it will give unbiased judgments [1]. In the current scenario, most of the advocates manually collect data and compare it to previous judicial statements and laws, then the final result is presented in front of the court. The judgment will be based on the manual analysis of the data presented. Day by day the data they need to refer to increases. As it is done manually, it will take a lot of time for each case to have its verdict. Due to corruption, there will bias judgement that will reduce the trust in the Indian judicial system. There are numerous cases reported all over India for producing one-sided judgment. And also, political parties in power will influence the judgement making it often wrong and illegal. To reduce that our method can be used in which CNN and NLP will be used for predicting the accurate verdict as per the law and transparent.

Our aim is to make computers to automatically extract information and derive a conclusion from it [11]. It will be very helpful for a layman to understand the approximate judgment of his case. Automatic analysis of the legal data presented will promote many courts to publish them online, to make it available to common people for future references [15][16].

In this study, the potential of using natural language analysis and automatic information extraction is addressed to facilitate statistical research in the legal domain [32]. More explicitly, the possibilities of Natural Language Processing techniques for automatically predicting judicial decisions of Indian Courts are established and discussed[8][9].

The main objective of this research work is to predict the verdict of the cases and to classify each case into bailable and non-bailable. For that most important part of this research work is to gather sufficient amounts of judicial data. After gathering, a natural language processing tool will extract keywords from it. And using neural networks for classifying the given text into several categories [4][6][19]. For example, for a criminal fact, a criminal charge may be classified as theft and rape from that the cases will be classified into bailable and non-bailable offence to further predict a judicial decision according to data given. Not saying that the exact judgment will get, but will get an approximate judgment of with least significant error [17].

II. BACKGROUND

Artificial intelligence and law are mixed way back in 1970, which lead to the introduction of the first lawyer robot. At that time people found out the reasoning process of humans can be planted into a robot in the form of program code. By statistical analysis the machine found out evidence and strategies just like a human would process a legal case or just like a lawyer.

The first International Conference on Artificial Intelligence and Law took place in 1897, at Northeastern University in Boston, USA. Eventually, it led to International Artificial Intelligence and Law being established in 1991 to increase the interest in the research and application in the interdisciplinary field of artificial intelligence and law [27].

There was so much research based on finding the word frequency in the text and the author used 16 judgments of the German Federal Constitutional Court on European Integration. The author found out both Word score and Wordfish can generate judicial positions according to the repetition of words in the text, he was able to find out the characteristics of the text[25]. Innumerable amounts of studies are there based on text mining arguments of legal cases. Able to identify arguments by automatically analyzing legal text and from that extracted data the decision can be predicted[26][31]. Natural language processing is one of the most common applications of text classification, it will automatically categorize mail, span recognitions, sentiment classification. The new coming data is labelled based on the training set of labelled data. The processing of text categorization divides into text pre-processing, text feature extraction, and classification model construction. And it is used for text extraction, word segmentation, and so on. There is research-based on text mining of legal cases.

International courts use automatic analysis of legal data and can be used for predicting court decisions. However, due to the huge amount of data accuracy is low. And a comparatively small number of studies have used machine learning methods for the analysis of the cases.US were the first to use this technique to predict the court decision or to find out the polling behavior of judge by making using of machine learning it was able to predict the decision of the court [24]. Their model achieved an accuracy of 70% at the case outcome level and 71% at the justice vote level. Outside the US only a few courts have used machine learning to predict the decision. A French court of cassation used machine learning for predicting the verdict. They aimed to predict the law area of a case and court ruling. Their model achieved an accuracy of over 92%. [23]

a) The Application of CNN in NLP

It is proven that CNN has significance in the image processing domain by classifying and identifying the object, but it is highly feasible for using CNN in NLP domains. By using NLP robot are becoming more and more close to human, understanding human characteristics from the feedback and processing it with a neural network would give the robot the intelligence for self-thinking. By implementing CNN in text processing, the entire text is represented by a vector. [21]

The input of NLP tasks is a document which is represented in a matrix and a convolutional layer will identify the pattern and information is broken homogeneously in each level. The convolutional neural network is similar to the human brain, that's why a convolutional neural network is more capable than standard machine learning. That's why CNN had made great progress in NLP such as sentiment analysis, translation, text prediction.

As the India Court is focused on, as it is having the biggest judicial laws. It is a tough challenge for us to integrate this much data and to predict the verdict which will satisfy the rules and regulations on the Indian constitution. Therefore, our research is started by using NLP and CNN methods using all the available data and exploring how can gradually improve accuracy on them so that in future 100% accurate decisions will be able to be predicted.

III. DATA COLLECTION

The publicly available data published by the Courts of India is being used. The dataset for this Experiment covers more than 200 criminal judicial records and 200 civil judicial records. Both are published by courts of India and it is available on their website.[15][16]

The laws, accusation and the penalties involved by violating each section is taken from the Indian constitution.

The entire dataset involves the violation of 26 variety of charges in the constitution. In these 26 charges, labelled and classified into a bailable and non-bailable offence using convolutional neural network, data is given to 2 layered neural networks and it will classify on the bases of charges, by identifying charges violated in that document. And then this charge is separated into bailable and non-bailable. Table 1 shows the data classification ratio of the section referred for the study. Table 2 shows 5 and this is presented below in Table 1. Hence the result shows there are 26 different varieties of a section where 39% is bailable and 61% is non-bailable. Table 2 represents some examples of bailable and non-bailable cases along with its section number. For each section or charges, there will be specific keywords that denote them. Classifying the legal data into bailable and non-bailable is part of the final process and done along with verdict prediction, its result is shown in Table I.

TABLE 1
Data Classification ratio of cases by ANN

Bailable	0.39
Non-Bailable	0.61

TABLE 2
Example of bailable and non-bailable cases showing section number, title and bailability.

Sections	Title	Bailable	Non-Bailable
121	Waging or attempting to wage war or abetting of war against the govt. of India	0	1
24 A	Conduct or speech inciting people to rebel against the authority of a state	0	1
172	Absconding(escape) to avoid service	0	1

232	Counterfeiting Indian coin	0	1
238	Export or import counterfeiting Indian coins	0	1
143	Being a member of an unlawful Assembly	1	0
148	Rioting, armed with a deadly weapon	1	0
171F	Bribery in relation to elections	1	0
193	Giving or fabricating false evidence in a judicial proceeding	1	0
273	Selling any food or drink as food and drink, knowing the same to be noxious	1	0

IV. FEATURE EXTRACTION OF LEGAL TEXT USING BAG OF WORDS

Legal data is very huge and it has some many unnecessary information, so it is a necessity to use a natural language processing toolbox to obtain required data (keywords) from it. The bag of words models used in our experiment, which will learn vocabulary from all the documents and then models each document by counting the number of times each word appears. By using the Bag of Words algorithm, the required keywords can get from the document at ease. In this model, a text is represented as the bag of its words, ignoring grammar and arrangement of the words but keeping multiplicity. The Bag of Words methods are used when needed to find the frequency of each word in the context, and by plotting it, will be able to identify the keywords in that case and these keywords will be the features of that case.

For Example,

Data: Madras High Court

Lakshmiammal vs Samiappa Goundar And Ors. on 18 September 1967“ the accused were armed with deadly weapons like knife, hammer, crowbar and spade; however, he held that the common object of the unlawful assembly was to break open the pipe and cause damage to-P.W. 1 and not to cause a threat or bodily injury to P.W. 1 and further held that though the accused were armed with weapons of the nature mentioned above, they cannot be said to be deadly weapons as they were not intended to be used as weapons of offences and so the accused have not committed the offence of rioting much less rioting with deadly weapons.”[17]

Our Vocabulary is as follow:

{1: accused,2: armed,3: deadly weapons,4: unlawful assembly,5: rioting}

As per table 2, it is a clear violation of IPC Section 148. This example shows how the Bag of Words will take the frequency of the words. But going deep into it will be cancelling and filtering each word so that can get exact words that convey the meaning of the passage.

After extracting the keywords from the text, keywords vs case number 2-dimensional array is created. After which, using CNN, the cases can be classified.

V. CONVOLUTIONAL NEURAL NETWORK FOR LEGAL TEXT CLASSIFICATION

The main idea of this study is to create a system, which can automatically classify into bailable or a non-bailable case and to predict the verdict of the case. For this task, needed to employ Convolutional Neural Network (CNN) for the classification because the convolutional neural network was originally designed to perform deep learning task, they use the concept of a “convolution”, a sliding window or “filter” that passes over the array of input, identifying important features and analyzing them one at a time then reducing them down to their essential characteristics and repeating the process until got the final product. Also experimented with multiple configurations of convolutional layers including 2 and 4. From these experiments, our best performing CNN model consisted of a configuration of two sets of two convolutional layers with each pair followed by a max-pooling layer. The number of neurons in the output layer of the model is the number of categories, and each neuron’s output is the probability of each accusation as shown in equation 1.

$$(f * g) = \int_{-\infty}^{\infty} f(r)g(t-r)\,dr \qquad \ldots(1)$$

a) Two-dimensional Convolution Layer

Matrix H = = {h1, h2, h3, h4….h_h}, H ∈ $R^{h\times d^{\omega}}$ where d^{ω} is the size of word embeddings. It will extract the features over H convolutional operation. Adding 2D filters were m ∈ $R^{k\times d}$, by designing a window for ‘k’ words and’ features.

The generated window of vectors: $H_{i:i+k-1,j:j+d-1}$

Range of i:1 to $l-k+1$ and j:1 to $d^{\omega}-d+1$

This filter is applied to each convolutional layer.

b) Two-dimensional Max Pooling Layer

2D Max pooling is done for fixed-length vector. h ∈ R and fixed length of h is $[l-k+1/p1] * [\,d^{\omega}-d+1/p2]$.

It applied to each window to extract the maximum value of that vector.

c) Output layer

For classifying the given legal text, output ‘h’ is passed through the SoftMax classifier layer.

Fig. 1 shows the proposed CNN model for classification and prediction. In this model, a fixed-length 256 is used as per the max-pooling equation in the 1st Convolutional Layer with a filter size of 5 across all convolutional layers. The convolutional layer would be extracting the important features present in the input vector and analyzing them one at a time then reducing them down to their essential characteristics.

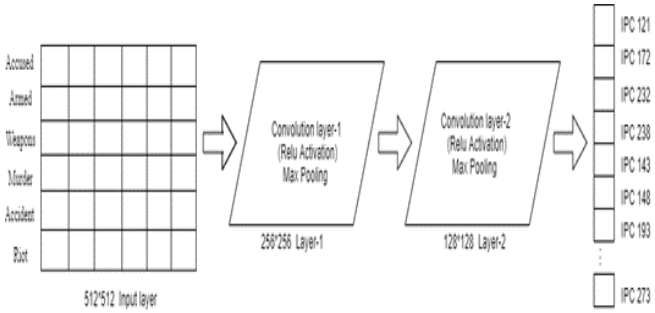

Fig. 1 CNN Model for prediction

After the second max pooling function, a dropout function is applied to help to prevent overfitting which would reduce the error so that accurate decision will get. In our model, a dropout rate of 0.5 is used. A fully connected layer is appended and then applied with a length of 128 followed by a second dropout function. This is followed by a dense layer with a size of 26 to represent the number of classification classes with a SoftMax function determining the output. The last layer represents 26 varieties of charges which are taken from the constitution of India. After identifying the charges, will be able to separate it as bailable or non-bailable cases. And for each charge, there is a separate verdict that can be mapped to it. Verdict for violation of more than one charge is given by mixing the verdicts of those two charges.

Our model is evaluated by using k-fold cross-validation. For that, the entire training data available for our model is taken to learn the characteristics of the cases which have given and split this into k distinct set. In our study k=4, fourfold cross-validation will be performed and train and test the model 4 times. i.e., one part is taken out and train the model using the remaining part of this set. Each time 75% of data is allotted for training and 25% is used for testing it. This way the model is more likely to perform better for unseen cases, its validation accuracy for each fold is shown in fig.6. In Fig.3 and Fig.4 show the graphical representation of the accuracy of our model. In this, by testing the model with our testing data accuracy of 85% will be got in our model.

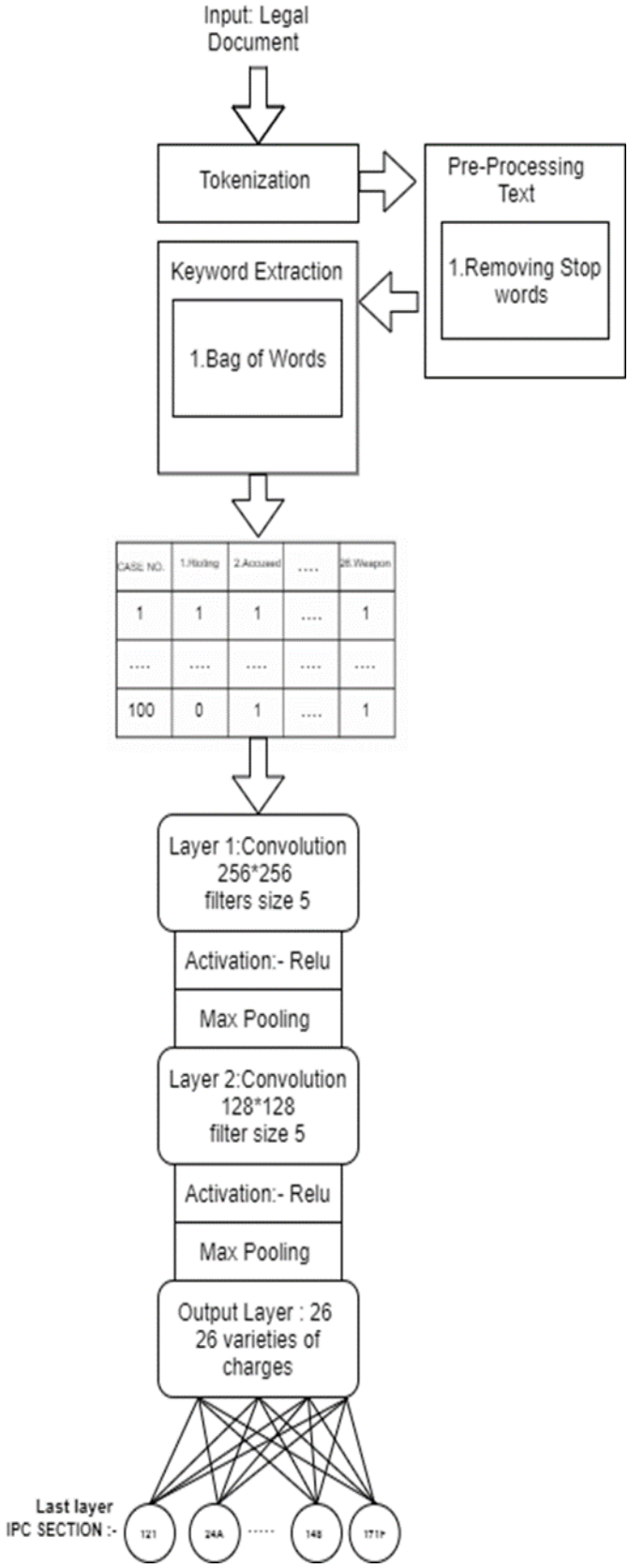

Fig.2 Detailed CNN Model with NLP for extraction of keywords and classification of law text.

VI. CNN MODEL ANALYSIS FOR VERDICT PREDICTION

This section describes how the study has been conducted. Fig.2 shows the frame diagram of a detailed CNN Model with NLP showing various stages for extraction of keywords and classification of law text in the process of predicting. Words from the judicial data will be extracted by using the Bag of Words method. Words segmentation and feature extraction from data is done by using the Bag of Words. And it is given the CNN model for the classification.

By extracting words, case facts can get. Due to the large text, a word segmentation tool is used to segment words. Due to data redundancy afterword segmentation, feature selection is required. By deleting punctuation and modal particles in content, a word list fit for text classification will get. As it contains 400 different cases 300 cases are used to train the classification model. A 2 -D array is made with keywords and cases that is the initial layer that is fed to the CNN model for classification. It consists of 2 layers of a convolutional network. Using that 300 cases a model is created to classify the cases into given 26 charges. For testing, the remaining 100 cases is used and got 85% accuracy in classifying new unseen data. Fig 6 shows the process of predicting. Word segmentation and feature extraction from data. And it is given the CNN model for the classification.

As can be seen from figure 1, the result is a vector. The dataset involves a violation of 26 charges, so the label dimension is 26 after Boolean encoding. For example, the charge in a case fact is [robbery, bribery,…], digitized to [1,2,7] and transformed into a vector [0,1,1,0,0,0,0,1,0,...,0].The transformed vector depends upon each case which gave as input to the convolutional neural network.

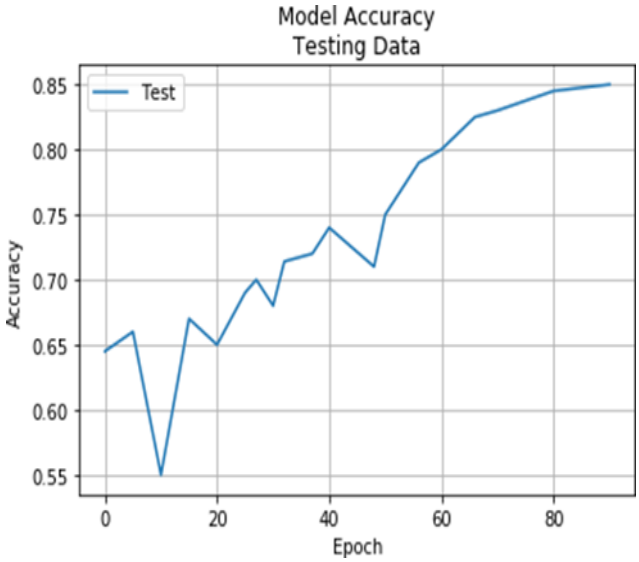

Fig. 3. Accuracy Plot of CNN Model during training

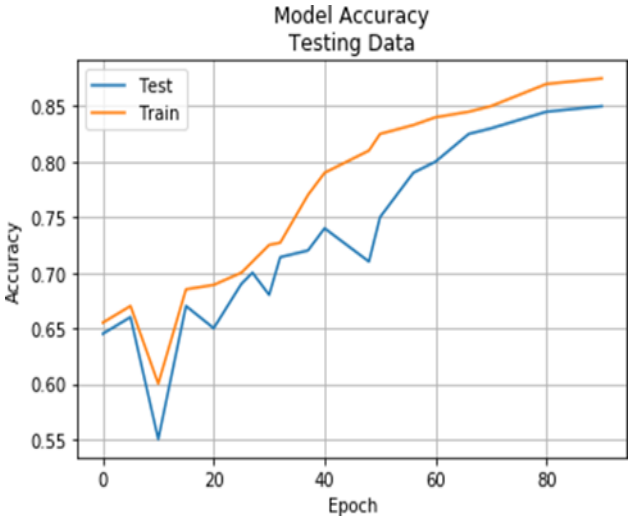

Fig. 4. Training and Testing accuracy plot vs epoch

According to charge in the cases, it is classified into bailable and non-bailable. And accusation label contains verdicts and penalties for each charge so mapping into it, the final verdict of the case will get. The result for 6 different cases is shown in table 3.

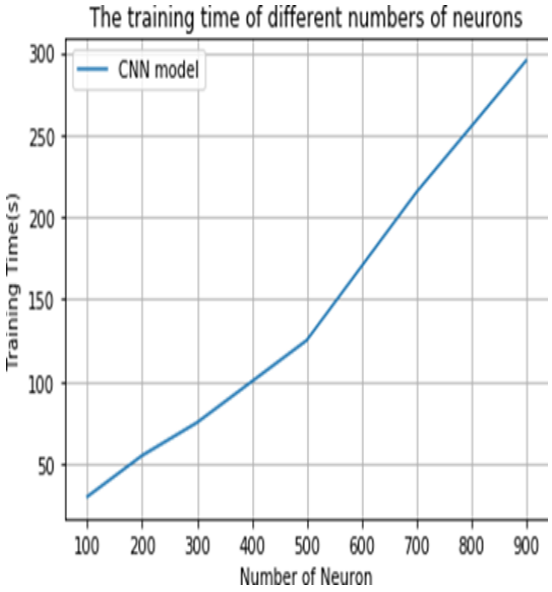

Fig.5. The training time of a different number of neurons.

Folds	Accuracy
Folds-1	84.57
Folds-2	85.5
Folds-3	85.4
Folds-4	84.6
Average	85.01

Fig.6.Validation Table of our model

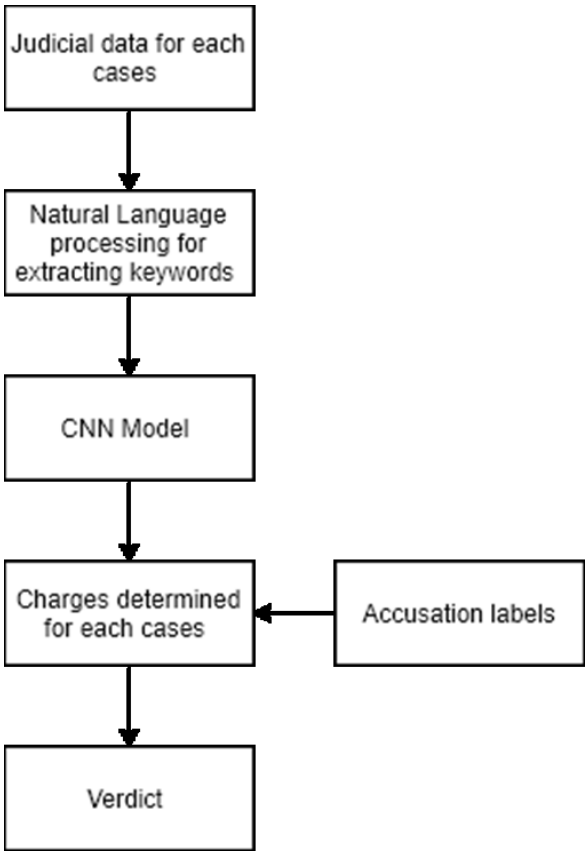

Fig.7. The flow chart of the process flow of verdict prediction

TABLE 3

Cases Number/ Label: -	Bailable	Non-bailable	Verdict
1	1	0	"IPC Section 143 whoever is a member of an unlawful assembly shall be punished with imprisonment of either description for a term which may extend to six months, or with fine, or with both".
2	1	0	"IPC Section 148 whoever is guilty of rioting, being armed with a deadly weapon or with anything which, used as a weapon of offence, is likely to cause death, shall be punished with imprisonment of either description for a term
			which may extend to three years, or with fine, or with both."
3	0	1	"IPC Section 121 Imprisonment for life up to 10 years along with fine."
4	1	0	"IPC Section 272 Imprisonment for 6 months with a fine of INR 1000."
5	0	1	"IPC Section 172 imprisonment for 1 month or fine with INR 1000."
6	0	1	"IPC Section 255 Imprisonment for 1 year with the fine."

Legal data from that legal data words are extracted and that extracted words are the same as input keywords. And the keyword-containing array is given to CNN and at the end, it gets classified into any of the 26 charges. And each charge contains its own IPC section. According to IPC section it gets further classified into bailable (yes in this case) or non-bailable. Based on the keywords the required IPC section will be predicted. In this case, the actual and predicted sections are the same.

False Data prediction.

Scenario 2: Classified as a violation of IPC 304A as shown in table 4.

TABLE 4

Verdict prediction used cases of false prediction using CNN Model.

Cases Number/Label: -	Bailable	Non-bailable	Original Verdict	Verdict Obtained

978-1-7281-5821-1/20/$31.00 ©2020 IEEE

51	1	0	"IPC Section *304A* sentenced to undergo Rigorous Imprisonment for six months IPC Section *279* Rigorous Imprisonment for two years and fine of Rs.500/- and in default of payment of fine to undergo...to Rigorous Imprisonment for one year"	"IPC Section 304A sent enced to undergo Rigorous Imprisonment for six months "

Verdict prediction used case examples and true classification of actual and predicted verdict cases using CNN Model.

Table 3 shows the prediction of verdicts and further classified into bailable and non-bailable. Our research dealt with only 26 charges and each legal data only contain a violation of a Data Analysis for True classification and prediction using case 1 in table 3.

Scenario 1: Classified as a violation of IPC 143

Keywords Extracted from data: -Accused, Armed, Deadly weapons, Unlawful Assembly, Rioting.

Verdict Obtained: - "IPC Section 143 whoever is a member of an unlawful assembly shall be punished with imprisonment of either description for a term which may extend to six months, or with fine, or with both".

Actual verdict: - "IPC Section 143 whoever is a member of an unlawful assembly shall be punished with imprisonment of either description for a term which may extend to six months, or with fine, or with both".ny of the 26 charges

Keywords Extracted from data: - Death, Negligence, Accused, Accident. Verdict Obtained: - "IPC Section 304A sentenced to undergo Rigorous Imprisonment for six months"

Actual Verdict: - "IPC Section 304A, 337 and 338 undergo violated. So, which deals with the death of a person by any rash or negligent act and leads to imprisonment up to 2 years. Both cases of accidents caused due to rash and negligent motor vehicle driving and also medical negligence leading to the death of a patient.

In this case, it combined the verdicts of two IPC sections but the actual verdict is different from it. So, increasing charges in a single case will give us undesirable judgments.Two different cases, one successful and unsuccessful case prediction was analyzed and that give an idea of what our model is doing by extracting keywords. Failure prediction due to fault in the extraction of keywords and lead to the wrong classification that's why got a wrong verdict.

CONCLUSION

For years the Indian judicial system depended on manual collecting and analysis of the evidence or data for judgements, but by introducing our CNN model, accurate and unbiased judgements can be given out in a much lesser time period. Moreover, it will be a great help to advocate less corruption and unbiased judgements. Our study depicts the importance of NLP and Convolutional Neural Network in judicial decision making by extracting words from the legal data and uses CNN to predict the verdict of the cases with 85% accuracy. According to the result, got the model works more efficiently and the case contains only one charge. As the number of charges increases the accuracy to predict verdict decreases.

In the future, looking forward to making a model which will work effectively with a greater number of cases.

REFERENCES

[1] W.Zhou, "On the artificial intelligence of judicial adjudication," in Wuhan University, 2018, pp. 1–60.
[2] R. E. Susskind, "Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning," The modern law review, vol. 49, no. 2, pp. 168–194, 1986.
[3] P. N. Gray, "The ontology of legal possibilities and legal potentialities." in LOAIT, 2007, pp. 7–23.
[4] R. Nanda, K. J. Adebayo, L. Di Caro, G. Boella, and L. Robaldo, "Legal information retrieval using topic clustering and neural networks." in COLIEE@ ICAIL, 2017, pp. 68–78.
[5] M. Grabmair and K. D. Ashley, "Facilitating case comparison using value judgments and intermediate legal concepts," in Proceedings of the 13th international conference on Artificial intelligence and law. ACM, 2011, pp. 161–170.
[6] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.
[7] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in International conference on machine learning, 2014, pp.1188–1196
[8] G. Greenleaf, "Legal expert systems-robot lawyers?(an introduction to knowledge-based applications to law)," 1989.
[9] T. Bench-Capon, M. Araszkiewicz, K. Ashley, K. Atkinson, F. Bex, F. Borges, D. Bourcier, P. Bourgine, J. G. Conrad, E. Francesconi et al., "A history of ai and law in 50 papers: 25 years of the international conference on ai and law," Artificial Intelligence and Law, vol. 20, no. 3,pp. 215– 319, 2012
[10] Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.218
[11] Dani Yogatama and Noah A. Smith, (2014), Making the Most of Bag of Words: Sentence Regularization with Alternating Direction Method of Multipliers, Proceedings of the 31st International Conference on Machine Learning,Beijing, China, JMLR: W&CP, volume 32
[12] Cordelia Schmid, Bag-of-features for category classification, www.cs.umd.edu/~djacobs/CMSC426/BagofWords.pdf
[13] Jialu Liu, Image Retrieval based on Bag-of-Words model, jialu.cs.illinois.edu/technical_notes/CBIR_Bag of Words.pdf
[14] Hiroshi Shimodaira, (2015), Text Classification using Naive Bayes, Learning and Data Note 7, Informatics 2B
[15] https://lawrato.com/indian-kanoon/ipc
[16] https://devgan.in/ipc
[17] https://indiankanoon.org/doc/840171/
[18] Pedro Domingos, A Few Useful Things to Know about MachineLearning,https://homes.cs.washington.edu/~pedrod/p apers/cacm12.pdf
[19] Joo-Chang Kim and Kyungyong Chung. "Neural-Network based Adaptive context prediction model for ambient intelligence" in Journal of Ambient Intelligence and Humanized Computing(2020).
[20] Dong Liu,Qirong Mao & Zhiyong Wang."Keyword Retrieving in continuous speech using temporal classification" in Journal of Ambient Intelligence and Humanized Computing(2020).
[21] S. Sandhiya & U. Palani,"An Effective Disease predicition system using incremental feature selection and temporal convolutional neural network"in in Journal of Ambient Intelligence and Humanized Computing(2020).
[22] Fen Zhao,Penghua Li ,Yuanyuan Li, Jie Hou and Yinguo Li , "Semi-Supervised Convolutional Neural Network for Law Advice Online"
[23] Sulea OM, Zampieri M, Vela M, van Genabith J.Predicting the law area and decisions of French Supreme Court cases. arXiv preprint
[24] Katz DM, Bommarito MJ II, Blackman J (2017) A general approach for predicting the behavior of the Supreme Court of the United States.
[25] http://www.wordfish.org.
[26] Mochales R, Moens MF (2008) Study on the structure of argumentation in case law. In: Proceedings of the 2008 conference on legal knowledge and information systems.
[27] Edwina L. Risslandt, "Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning."

[28] A. S. Kumar and Jyothisha J. Nair, "Pair wise training for stacked convolutional autoencoders using small scale images", Journal of Intelligent and Fuzzy Systems, vol. 36, pp. 1987-1995

[29] Joseph, Anju; Vineetha, K. V.; Kurup, Dhanesh G.; Mini, R. "A neural network based overvoltage prediction system for long cable issue" in Intelligent Computing, Information and Control Systems, Springer International Publishing, Cham (2019)

[30] Thangavel, Senthil; S, Rajendrakumar; Rajevan, T "Deep learning based emotion analysis approach for strengthening teaching learning process of tribal school" in Journal of Advanced Research in Dynamical and Control Systems, Volume 11, p.621-635 (2019).

[31] Veena, G., Gupta, D., Anil, A., Akhil, S.,"An Ontology Driven Question Answering System for Legal Documents ",2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2019.

[32] L R Pillai,Veena G and Gupta D, "A Combined Approach Using Semantic Role Labelling and Word Sense Disambiguation for Question Generation and Answer Extraction" in Second International Conference on Advances in Electronics Computer and Communication(ICAECC),Bangalore,India,2018