2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI) | 978-1-6654-3875-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/IRI51335.2021.00022

Identifying the Severity of Road Accident Impact on Traffic Flow by Ensemble Model

Zinat Ara
Department of Information Sciences and Technology
George Mason University
Fairfax, VA, USA
zara@gmu.edu

Mahdi Hashemi
Department of Information Sciences and Technology
George Mason University
Fairfax, VA, USA
mhashem2@gmu.edu, ORCID: 0000-0003-0212-0228

***Abstract*— Road accidents are one of the prime reasons for traffic congestions and delays. Predicting the severity of traffic congestions caused by road accidents can help to prepare travelers and traffic managers for the delays and re-routings. This study constructs a Stacked Ensembled Classifier model (SEC) by combining three ensemble machine learning algorithms Random Forest (RF), Extreme Gradient Boosting (XGB) and Gradient Boosting Machines (GBM). Severity of the impact of a road accident on the current traffic flow is predicted based on time, weather, distance, location, road conditions, traffic signal and other points of interests. These three ensemble models are chosen because they achieved higher accuracies than other baseline models. The results indicate that our stacked classifier outperforms individual performance of the ensemble algorithms by up to 3.2%. Additionally, the average daily traffic count on each road segment, whose data is obtained from Virginia department of transportation, is introduced here as a new input feature. Inclusion of this feature improves the prediction accuracy up to 4.6%.**

Keywords—traffic flow, ensemble model, machine learning, average daily traffic

I. Introduction

In the modern era, intelligent transport system has become a major part in urbanization. Researchers have designed and modeled numerous smart applications to tackle issues such as addressing safety, convenience, performance and reliability. Safety concerns arise when latest technology fail to provide useful solutions to avoid road accidents and its impacts. Researchers have identified some influential impact factors for predicting traffic accidents and proposed a multitude of solutions to detect its severity [1, 2, 3, 4, 5]. Also a variety of research have been conducted for predicting traffic flow using machine learning [6, 7, 8, 9]. When an accident occurs, it impacts road traffic by creating a delay, and forecasting the severity of the impact to travelers can help them decide their next step. This information is useful for a smarter navigation system. Travelers and traffic managers will be able to choose an alternate route and avoid unwanted congestions. Although there exists many research to observe the effect of traffic congestion created due to road accidents in other disciplines [10, 11, 12], fewer instances of solutions using applied machine learning are found to address this issue. This study determines the impact of road crash incident on the traffic flow for a particular road using an ensembled stacked classifier. Stacking itself is an ensemble machine learning algorithm that learns how to best combine the predictions from multiple well-performing machine learning models. The goal is to predict severity level of the accident impacts (i.e. short or long delay) on the current traffic. The results from our proposed model is compared and analyzed with other baseline models. The proposed method also includes the process of extracting some influential factors for accident analysis from two real world datasets.

The output is severity which is the impact of accident over traffic flow, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact (i.e., long delay). While the dataset used does not provide a translation of severity level to exact amount of traffic delay (in unit of time), an estimation can be calculated with the help of other datasets. This paper combined three ensemble machine learning algorithms Random Forest (RF), Extreme Gradient Boosting (XGB) and Gradient Boosting Machines (GBM) to create a stacked generalization model which harnesses the capabilities of these models and make predictions that have better performance than any single model in the ensemble.

II. Related Work

Machine learning (ML) and prediction models are used by researchers to make transportation systems more intelligent [13, 14, 15, 16, 17, 18, 19, 20]. A wide spectrum of research has been done towards predicting the road accidents and traffic congestion. Majority of these past research works have predicted accident severity based on injury level or fatality rate using machine learning techniques [1, 2, 3, 4]. Karlaftis and Vlahogianni [21] developed a comprehensive review on differences, similarities and insights between the statistical or computational methods versus neural networks in transportation research. Both of these approaches have advantages and disadvantages depending on the problem formulation and designing proper prediction model. Iranitalab and Khattak [22] compared four statistical and machine learning methods which are multinomial logit, nearest neighbor classification, support vector machines and random forests to predict traffic crash severity. Its prediction rates show that k-nearest neighbor had the best performance in severe and minor crashes and multinomial logit shows weakest estimates. Nguyen et al. [23] applied ML techniques to classify traffic incident severity from the textual logs provided by the transport management center. They considered the factors causing the accident and implemented Naive Bayes, support vector machines and decision tree models where decision tree-based model

978-1-6654-3875-9/21/$31.00 ©2021 IEEE
DOI 10.1109/IRI51335.2021.00022

outperforms the others. Deep learning based models are implemented in accident analysis as well. Dong et al. [3] developed a deep learning model with two modules, an unsupervised module for feature learning and supervised module for predicting crash severity.

Navigation systems can compute more efficient routes by knowing traffic patterns in advance which in return saves time for the travelers [3]. Lesham and Ritov [7] proposed a method of forecasting traffic congestion combining random forest algorithm into Adaboost which gives promising results on both simulations and real data. Alajali et al. [8] presented some ensemble decision tree models to predict traffic flow in the road intersections by introducing extra data sources along with road traffic volume data into the prediction model.

Retallack and Ostendorf [10] studied the correlation between traffic congestion and number of traffic accidents along with their risk factor. Bei Pan et al. [24] addressed the problem of predicting and quantifying the impact of non-recurring traffic incidents including accidents, weather hazard, road construction or work zone closures. Nowakowska [25] used logistic regression in the form of the proportional odds ratio model and continuation ratio logits to identify the features that have a statistically significant influence on accident severity. Though some of these works come from different areas of studies, they mostly focus on theoretical analysis of the effect that traffic crashes have on traffic flow. The majority of the authors from previous works mainly applied available ML algorithms on their data with giving less focus on choosing the best train set for improving model accuracy. This paper proposes a model which combines three ensemble machine learning algorithms random forest, extreme gradient boosting and gradient boosting machine. Different machine learning models have different assumptions on same dataset to solve a prediction task. Stacking method is appropriate in this case to estimate prediction range using those models and it is designed to improve modelling performance. The improved model explores deeper into the impact of traffic accident on the traffic flow to extract any unknown patterns. Besides investigating better prediction model for crash impact severity classification, an important feature is also introduced strongly related to the output parameter. Adding this feature has improved the performance of all classification models that are used.

III. Data Description

Two real world traffic datasets are collected for this study. First is US Accidents, which is a countrywide traffic accident dataset. It consists of a total 3.5 million accident reports occurring in 49 states of the USA from February 2016 to June 2020. The data is continuously being collected using several data providers, including two APIs which provide streaming traffic event data [26, 27].

Second is average daily traffic volumes with vehicle classification data provided by Virginia State Department of Transportation (VDOT). The reason for choosing this dataset is it provides the Annual Average Daily Traffic (AADT) count which is an estimate of the average daily traffic along a defined segment of roadway. This value is an important indicator of the average traffic flow for a road segment. It is calculated from short term counts taken along the same road section which are then factored to produce the estimate of AADT [28].

Severity level will be predicted using this AADT as an additional input feature along with other influential factors such as weather, time, location, road conditions, distance which are provided in the US accident data. This severity denotes the impact of an accident on the traffic flow and it is classified as a number from 1 to 4, where 1 indicates the least impact on traffic and 4 indicates a significant impact (i.e., long delay)

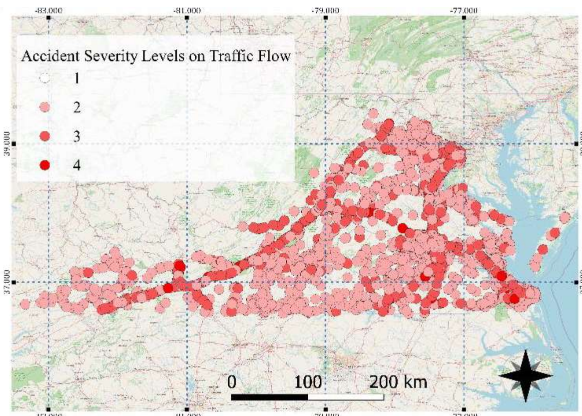

Figure 1 Accident locations with respective severity levels in Virginia state from January 2019 - June 2020

The experiment is run on Virginia state data. VA state accident records are filtered from January 2019 to 2020 June from the original source. The shape of extracted dataset is 15,500 records with 49 parameters. Figure 1 plots the accident occurrence spots along with its severity level measuring on latitude and longitude of each location.

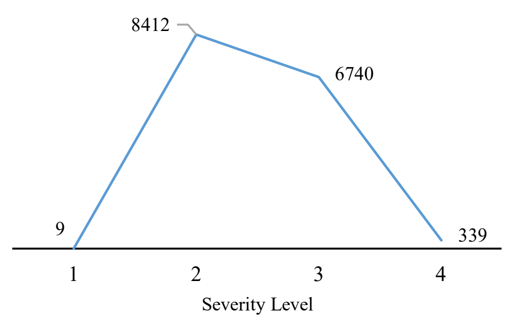

Figure 2 Number of accident records for each severity level in VA state between January 2019 to June 2020

From figure 2, it can be observed that there is a noticeable lack of values for severity level 1 and severity level 4. Severity 2 and 3 take up majority portion of the distribution styled.

While working with a big dataset, finding relevant features for ML models are necessary. Using all of the features may reduce the performance of the algorithm. Malin et al. [28] investigated the frequency and relative accident risk with different weather and road conditions. For this study 16 features are selected from US accident dataset which may have impact on road accidents. These features are related to weather, time, location and different road and traffic conditions and other point of interests (POI). They are further defined in Table 1.

US accident dataset captures all relevant information regarding an accident, but it does not provide any information regarding the road segment traffic count of the accident location. This information is obtained from VDOT site, the source of AADT count for each segment.

TABLE 1 DESCRIPTION OF SELECTED INPUT FEATURES USED FOR CLASSIFICATION

Input Features	Description
Severity	Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).
Distance	The length of the road extent affected by the accident.
City	Shows the city in address record.
County	Shows the county in address record.
Side	Shows the relative side of the street (Right/Left) in address record.
Temperature	Shows the temperature (in Fahrenheit).
Humidity	Shows the humidity (in percentage).
Visibility	Shows visibility (in miles).
Weather Condition	Shows the weather condition (rain, snow, thunderstorm, fog, etc.)
Amenity	A POI annotation which indicates presence of amenity in a nearby location.
Precipitation	Shows precipitation amount in inches if there is any.
Crossing	A POI annotation which indicates presence of crossing in a nearby location.
Traffic Signal	A POI annotation which indicates presence of traffic signal in a nearby location.
Hour	The hour of day when the accident occurred.
Weekday	Converting the start time to indicate the day of the week.
Time Duration (min)	Duration of the accident extracted from (End Time – Start Time).

IV. METHODOLOGY

Supervised learning is a type of machine learning approach where the machine is trained with some input variables (X) and the algorithm learns a mapping function Y = f(X) to deduce the output variables. The severity prediction problem addressed here is a classification task. Classification is the prediction of a categorial variable within a predefined vocabulary based on training examples. In the first three sections, tree based ensembled classification models are described. Our proposed model, the process of extracting AADT and its integration is discussed in later sections.

Severity prediction using above algorithms is done using historical data of 16 selected features from Table 1. A new feature AADT is included along with the existing ones to observe the outcome of the models. The end result for this additional input is shown in the result section. Final result is compared with some baseline models. Random forest model is used to rank the feature importance while the other models and RF itself were employed to analyze accident impact severity.

A. Gradient Boosting Machine (GBM)

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function [29]. Decision tree is used frequently to predict and rank different features in traffic safety studies [8, 30]. Several variants of decision tree models have been introduced, depending on the complexity of the operations allowed in the computation of a single comparison and the way of branching.

Gradient boosting is typically used with decision trees (especially CART trees) of a fixed size as base learners. For this special case, Friedman [29] proposes a modification to gradient boosting method which improves the quality of fit of each base learner.

Generic gradient boosting at the m-th step would fit a decision tree $h_m(x)$ to pseudo-residuals. Let J_m be the number of its leaves. The tree partitions the input space into J_m disjoint regions R_{1m} ,,,, $R_{j_m m}$ and predicts a constant value in each region. The output of $h_m(x)$ for input x can be written as the sum:

$$h_m(x) = \sum_{j=1}^{J_m} b_{jm} \, 1 \, R_{jm}(x) \tag{1}$$

where b_{jm} is the value predicted in the region R_{jm} . Then the coefficients b_{jm} are multiplied by some value γ_m chosen using line search so as to minimize the loss function, and the model is updated as follows:

$$F_m(x) = F_{m-1}(x)\gamma_m h_m(x),$$

$$\gamma_m = \arg min \sum_{i=1}^{n} L(y_{i,} F_{m-1}(x_i) + \gamma h_m(x_i)) \tag{2}$$

Friedman proposes to modify this algorithm so that it chooses a separate optimal value γ_{jm} for each of the tree's regions, instead of a single γ_m for the whole tree. This modified algorithm is known as TreeBoost [29].

In this paper GBM builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage 100 regression trees are fit on the multinomial deviance loss function.

B. Random Forest (RF)

Random Forest is a prediction method categorized as ensemble learning, which refers to methods that generate many classifiers and aggregate their results [31]. Breiman [32] proposed this method as a prediction tool consisting of a collection of tree-structured classifiers with independent identically distributed random vectors, while each tree casts a unit vote for the most popular class at input as in Figure 3. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. During the tree growing procedure, about one-third of the training data were left out from the training trees, which become the OOB (out-of-bag) data. The OOB data are utilized to achieve unbiased estimate of variable importance as trees are added to the forest [1]. This method has performed very well compared to many other commonly- used classifiers and is robust against overfitting [32]. Implementing this method requires determination of the models' parameters, including number of trees to grow and number of variables randomly sampled as candidates at each split [22].

This study has used RF to predict severity as one of the base level models and also to rank the importance of features which identifies the most influential factor in the model. Number of estimators have been set up to 100.

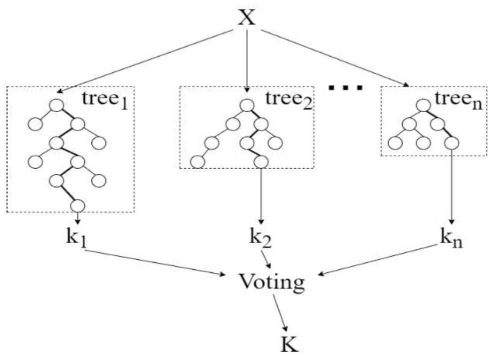

Figure 3 Structure of a random forest model

C. Extreme Gradient Boosting (XGB)

XGB is an integrated learning parallel processing algorithm based on tree structure [33], is nonparameterized and can deal with the complex nonlinear relationships between features. The idea of this algorithm is to continuously add trees, continuously perform feature splitting to grow a tree and each time to add a tree, needing to learn a new function to fit the last predicted residual [34]. To forecast a sample score, one should get K trees through a training sample dataset. This process corresponds to the characteristics of the sample, each characteristic will fall into a corresponding leaf node, and each leaf node will correspond to a score. Finally, we need to add each tree's corresponding score to obtain the predicted value of the sample. XGB's tree integration model is essentially a set of classification regression trees (CARTs). Each tree (CART) means a decision model $f(.)$. The XGB algorithm uses model integration; to be precise, it uses k decision tree models to output results based on input x_i and sums the K output results to obtain $\hat{y}$. The set of decision tree models is called F.

$$\hat{y} = \emptyset(x_i) = \sum_{k=1}^{K} f_k(x_i),$$

$$F = \{f(x) = w_{q(x)}\}, \qquad (q: \mathbb{R}^m \to T, w \in \mathbb{R}^T) \qquad (3)$$

where F is a set that incorporates k decision tree models, among them $F = f_1(.), f_2(.), f_3(.), \dots f_k(.)$: $f_k(.)$ represents the input/output function relationship of the kth regression tree corresponds to the structure q and leaf node weight w of the kth regression tree. w_i represents the score of the ith leaf node. Therefore, the tth iteration of predicted values can be displayed as:

$$\hat{y}^t = \sum_{k=1}^{t} f_k(x_i) = \hat{y}^{t-1} + f_t(x_i) \qquad (4)$$

The classification regression trees are generally arranged in a sequence, but the trees generated by the XGB training are arranged in a parallel manner.

D. Proposed method: Stacked Ensembled Classifier (SEC)

Stacking is a process where output from multiple machine learning models are used to generate a final prediction on the same dataset, like bagging and boosting. It is an ensemble machine learning approach that learns how to best combine the predictions from multiple well-performing machine learning models *[35]*.

The architecture of a stacking model involves two or more base models, often referred to as level-0 models, and a meta-model that combines the predictions of the base models, referred to as a level-1 model. First, we seek to find the most satisfactory single ML algorithms and ensemble models based on their performance. After observing the performance components of the stacked model are chosen based on their accuracies.

- Level-0 Models (Base-Models): Models are fit with the training data and their predictions are compiled. In this study random forest, XGB and GBM are used as base models. The reason for choosing these 3 is because they achieved higher accuracies compared to other ML algorithms for the stated problem.
- Level-1 Model (Meta-Model): Model that learns how to best combine the predictions of the base models. In our model XGB is used as meta model. XGB was chosen as the meta model for its faster performance and higher accuracy.

The meta-model trains using the predictions made by base models on out-of-sample data, which is, data not used to train the base models. These predictions, along with the expected outputs, provide the input and output pairs of the training dataset used to fit the meta-model. We have used the extreme gradient boosting technique for our meta model with the maximum depth of the tree set to 6. The loss function is defined as multi softmax and set to predict 4 classes. The outputs from the base models used as input to the meta-model may be probability values, probability like values, or class labels in the case of classification [35]. In our study we are predicting class label for severity.

The most common approach to preparing the training dataset for the meta-model is via k-fold cross-validation of the base models, where the out-of-fold predictions are used as the basis for the training dataset for the meta-model. Our model uses stratified k-fold 10 times with 5 folds for each repetition. Figure 4 illustrates the architecture of the proposed stacked ensembled model. The training data for XGB also include the inputs to the base models, e.g. input elements of the training data. This can provide an additional context to the meta-model as to how to best combine the predictions from the meta-model. Once the training dataset is prepared for the XGB, it can be trained in isolation on this dataset, and the base-models can be trained on the entire original training dataset.

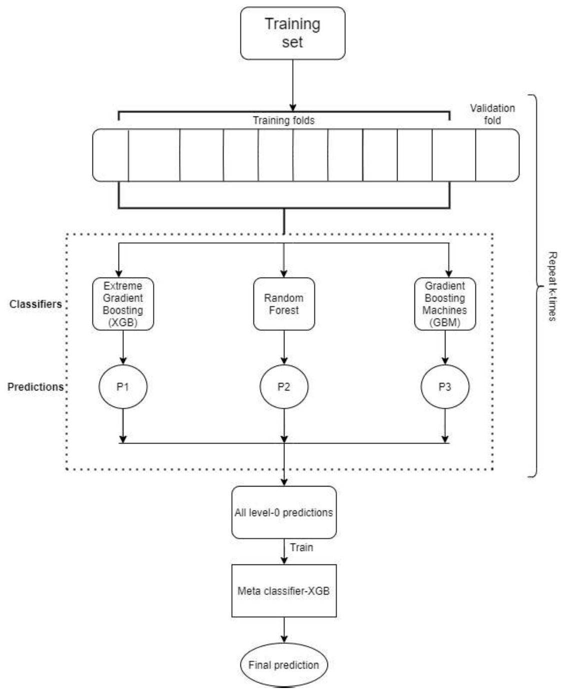

Figure 4 Architecture of Stacked Ensembled Classifier (SEC)

Given a set of observations $\varphi = \{x_i \in \mathbb{R}^n\}$ and a set of labels $N = \{y_i \in Y\}$ and a training set $DF = \{x_i, y_i\}_{i=1}^{m}$ as an input we want to solve the classification where our model learns a classifier f_t based on *DF*. Table 2 shows the algorithm of the stacked model with cross validation [36, 37] in details.

TABLE 2 ALGORITHM OF STACKING GENERALIZATION

Algorithm: Stacking with K-fold cross validation
Input: Training dataset $DF = \{x_i, y_i\}_{i=1}^{m}(x_i \in \mathbb{R}^n, y_i \in Y)$
Output: An ensemble classifier S
1. Step 1: Adopt cross validation approach in preparing a training set for level 1 classifier
2. Randomly split *DF* into *K* equal-size subsets: $DF = \{DF_1, DF_2, \ldots DF_k\}$
3. **for** $k \Leftarrow 1$ to K **do**
4. Learn level 0 classifiers
5. **for** $t \Leftarrow 1$ to T **do**
6. Learn a classifier f_{kt} from DF_k
7. **end for**
8. Construct a training set for level 1 classifier
9. **for** $x_i \in DF_k$ **do**
10. get a record $\{x_i' = \{f_{k1}(x_i), f_{k2}(x_i), \ldots, f_{kT}(x_i)\}$
11. **end for**
12. **end for**
13. Learn a level 1 classifier
14. Learn a new classifier f' from the collection of $\{x_i', y_i\}$
15. Re-learn level 0 classifiers
16. **for** $t \Leftarrow 1$ to T **do**
17. Learn a classifier f_t based on *DF*
18. **end for**
19. **return** $F(x) = f'(f_1(x), f_2(x), \ldots, f_T(x))$

The datasets are randomly divided into k sets. Each dataset is cross validated for a number of T times to learn a new classifier f_{kt}. For each fold, the process constructs a new training set for next level classifier. Using this classifier, each record x_i' is called in order to predict the label value.

E. AADT Extraction

US accident dataset provided us the latitude and longitude information of the accident occurrence spot. ARCGIS Rest API [38] is used to plot these accident spot on the map to find the exact road segment. The degree decimal format of the accident spot was converted into X Y coordinates for matching it with the input parameter. The input geometry takes an envelope consisting of $x_{max}, x_{min}, y_{max}, y_{min}$ coordinates which defines an area around the designated spot. The difference between the actual X/Y coordinate and max/min coordinate is identified as a deviation. Starting the deviation as 3, the value is gradually increased if no road segment is found within the accident area. Once the correct road segment is found, AADT value is extracted from the API output. The whole process of finding road segment and corresponding AADT was automated for 15500 accident records. Figure 5 demonstrates this process in a flow diagram.

Since the objective of this analysis is to identify the severity on traffic flow and the effects of variables causing accidents are not to be examined, to avoid overfitting, the input features selected in this study are based on previous crash severity literature and ratio of missing values in the dataset. Inclusion of variables that were reported not statistically significant in previous studies are disregarded. Features that would decrease the sample size due to missing values were also avoided. Pearson's correlation coefficient is used to measure the correlation among the features. Feature selection benefits the model by simplifying and shortening its training times and enhances generalization by reducing overfitting the method. Factors like time duration and hour calculation showed significant effect on previous studies [39, 10, 40, 41, 42]. These features are extracted using given information. Time duration feature is calculated by calculating the difference between start time and end time features, hour and weekday features are generated by converting start time parameter into hour and day of the week respectively [38]. As outliers are one of the main problems to build a better prediction model, we need to normalize the data to fit in the model and get satisfactory result. This was handled by replacing the outlier values with the median of the data.

V. RESULTS AND DISCUSSION

This section includes the detailed comparison of three individual algorithms with stacked ensembled classification (SEC) model and discusses the effect of the additional feature AADT. Pearson's correlation is one of the feature selection methods which is used in determining the linear relationship between variables.

Figure 6 plots the correlation coefficients matrix of the selected features including AADT. From the heatmap scale it is

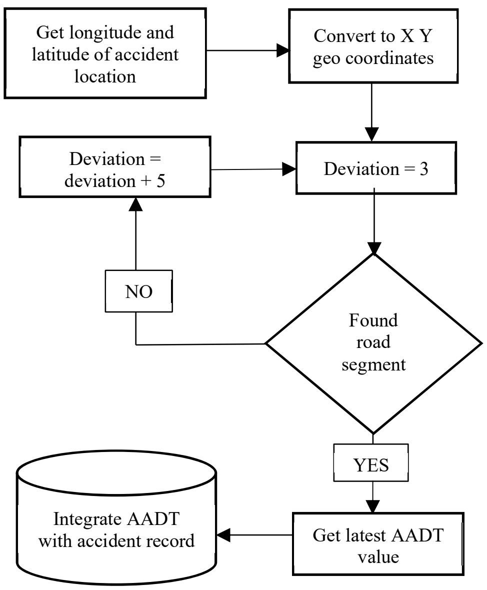

Figure 5 Flow diagram of AADT extraction process

clear that AADT is strongly related with severity. Other significant features such as hour, temperature, distance are identified to have positive relation with severity whereas traffic signal, crossing features are negatively correlated.

In multilabel classification, the function returns the subset accuracy. If the entire set of predicted labels for a sample strictly match with the true set of labels, then the subset accuracy is 1.0; otherwise it is 0.0. In general if $\hat{y}_i$ is the predicted value of the i-th sample and y_i is the corresponding true value, then the fraction of correct predictions over $n_{samples}$ is defined as:

$$accuracy\ (y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \qquad (5)$$

Figure 7 shows the accuracy comparison between algorithms with SEC before and after adding AADT. From the graph it can be observed that SEC produces higher accuracy score over the other techniques. Also, the inclusion of AADT while classifying, improves the result for all models. SEC produces an accuracy of 0.821 and 0.79. XGB performs closest to SEC with an accuracy score of 0.819 and 0.774 with and without AADT. Figure 8 depicts the comparison between the proposed method with other machine learning models and SEC outperforms all. The primary reason for choosing the ensembled models in SEC over the ML models is their better accuracies. Figure 9 shows the outperformance percentage of SEC over random forest, XGB and GBM with and without AADT. SEC shows the highest performance increase over GBM which is 2.1% with AADT and 3.2% without AADT. Random Forest and XGB get improvement of 2.2% and 1.2% respectively without AADT.

Figure 10 shows feature importance scores for top 10 features during prediction. AADT surpasses the other features.

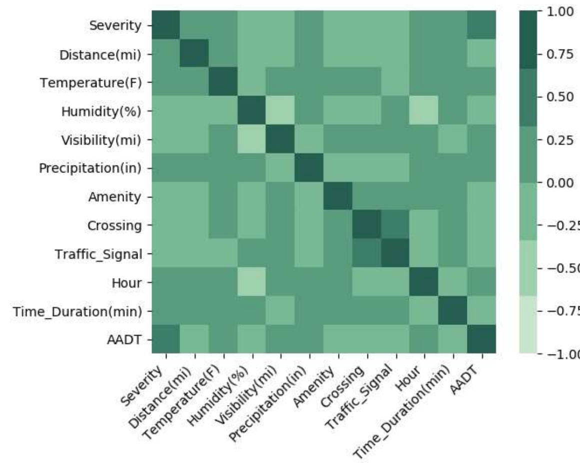

Figure 6 Correlation coefficients for selected features

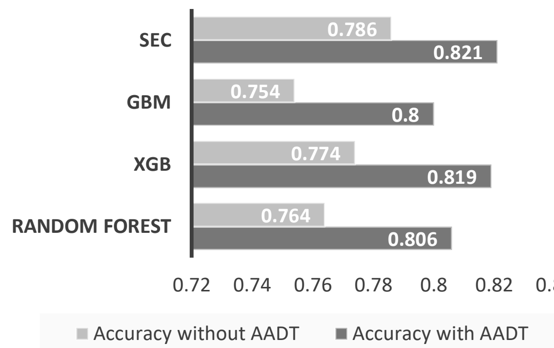

Figure 7 Comparing accuracy before and after adding AADT for each model

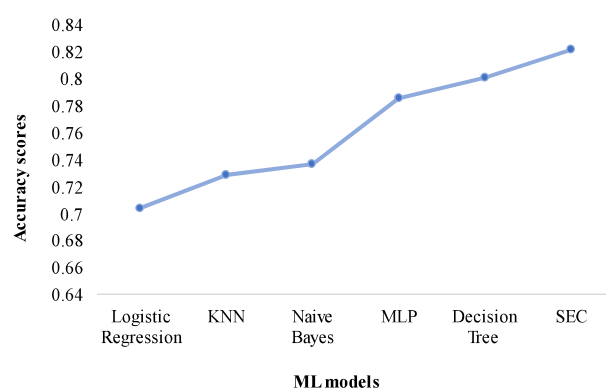

Figure 8 Accuracy comparison with other ML models and SEC (withAADT)

The next two most influential features are hour and time duration (min) which were extracted from the dataset features.

Precision-Recall is another useful measure of success of prediction when the classes are imbalanced. In information retrieval, precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. Precision P is defined as the number of true positives T_p over the number of true positives plus the number of false positives F_p

$$P = \frac{T_p}{T_p+F_p} \quad (6)$$

Recall R is defined as the number of true positives T_p over the number of true positives plus the number of false negatives F_n

$$R = \frac{T_p}{T_p+F_n} \quad (7)$$

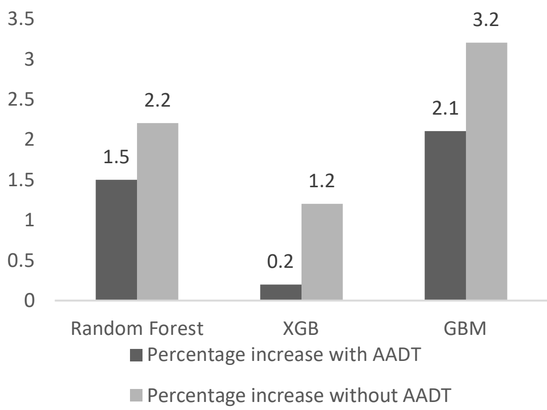

Figure 9 Accuracy increase in percentage for SEC model comparing with other models

These quantities are also related to the F_1 score, which is defined as the harmonic mean of precision and recall.

TABLE 3 CLASSIFICATION REPORT FOR ALL PREDICTION MODELS

Severity	Precision	Recall	F1-Score	Algorithm
1	1	0	0	**XGB**
2	0.84	0.87	0.86	
3	0.80	0.79	0.79	
4	1	0.26	0.41	
Macro Avg.	0.91	0.48	0.52	
Weighted Avg.	0.83	0.82	0.82	
1	1	0	0	**Random Forest**
2	0.84	0.88	0.86	
3	0.81	0.77	0.79	
4	0.47	0.23	0.30	
Macro Avg.	0.78	0.47	0.49	
Weighted Avg.	0.82	0.82	0.82	
1	1	0	0	**GBM**
2	0.81	0.90	0.85	
3	0.82	0.73	0.77	
4	0.80	0.13	0.22	
Macro Avg.	0.86	0.44	0.46	
Weighted Avg.	0.81	0.81	0.80	
1	1	0	0	**SEC**
2	0.86	0.86	0.86	
3	0.79	0.82	0.80	
4	0.62	0.26	0.36	
Macro Avg.	0.82	0.48	0.51	
Weighted Avg.	0.83	0.83	0.83	

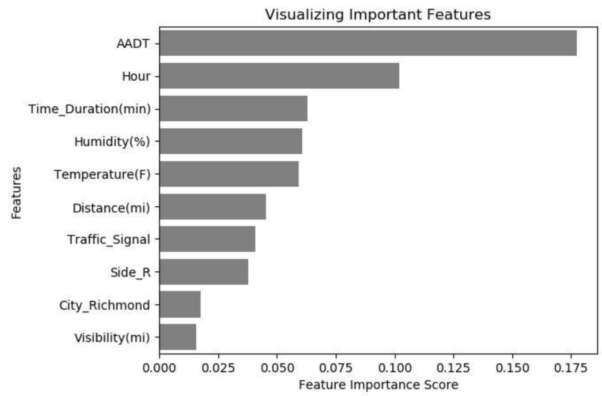

Figure 10 Comparison of feature importance scores during prediction

$$F_1 = 2\,\frac{P \times R}{P + R} \quad (8)$$

Table 3 shows the corresponding precision, recall and F1 scores for each algorithm considering AADT. The reported averages include macro average (averaging the unweighted mean per label) and weighted average (averaging the support-weighted mean per label). Considering the weighted average, random forest gives us the best output in precision and recall followed by decision tree. As our data was not equally balanced for all severity levels, for severity 1 and 4, recall and F1-scores are relatively lower for some of the algorithms. This results in lower macro average values for all three criteria, but the weighted averages exceed them by far.

VI. CONCLUSION AND FUTURE WORK

Road accidents can have significant effect on traffic congestion and for intelligent transportation system it is crucial to predict this impact with high accuracy. This information can be used to build smart application for warning people facilitating an intelligent navigation system. However, few studies have considered the relationship between the average traffic count with the regular traffic flow.

This study carried out detailed analyses, visualizations and classification of the accidents reported from entire Virginia state road network for the year 2019 to June 2020. We have investigated the effect of accident on traffic flow in terms of severity level. Higher the severity level, greater the impact will be and as the traffic delay is positively corelated to severity. To get the best possible outcome different tree based ensembled models are utilized to generate a new prediction model by stacking approach. The advantage of stacking algorithm is, it combines the power of multiple ML techniques and can provide better output. Our proposed Stacked Ensembled Classifier (SEC) combines RF, XGB and GBM and outperforms all models. Choosing the right set of parameters is important to accurately identify and we have managed to import a new attribute from another historical dataset. This addition has proved to be the most important influencer among the other explanatory variables typically used in the crash injury severity studies.

In future, this study can be extended by exploring regression and deep learning based models to improve the performance.

Several other datasets can be investigated to find new set of potential factors that can influence our model. Predicting the amount of traffic delay time precisely based on severity level can also be a new research topic.

REFERENCES

[1] R. Yu and M. Abdel-Aty, "Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data," *Safety Science,* vol. 63, pp. 50-56, 2014.

[2] G. Liang, "Automatic Traffic Accident Detection Based on theInternet of Things and Support Vector Machine," *International Journal of Smart Home,* vol. 9, no. 4, pp. 97-106, 2015.

[3] C. Dong, C. Shao, J. Li and Z. Xiong, "An Improved Deep Learning Model for Traffic Crash Prediction," *Journal of Advanced Transportation,* vol. 2018, p. 13, 2018.

[4] N. Dogru and A. Subasi, "Traffic Accident Detection Using Random Forest Classifier," in *15th Learning and Technology Conference (L&T)*, Jeddah, 2018.

[5] F. Bhatti, M. Ali Shah, C. Maple and S. Ul Islam, "A Novel Internet of Things-Enabled AccidentDetection and Reporting System for SmartCity Environments," *sensors,* vol. 19, 2019.

[6] S. Kwoczek, S. Di Martino and W. Nejdl, "Predicting and visualizing traffic congestion in the presence of planned special events," *Journal of Visual Languages and Computing,* vol. 25, pp. 973-980, 2014.

[7] G. Leshem and Y. Ritov, "Traffic Flow Prediction using Adaboost Algorithm with Random Forests as a Weak Learner," *International Journal of Mathematical and Computational Sciences,* vol. 1, no. 1, 2007.

[8] W. Alajali, W. Zhou, S. Wen and Y. Wang, "Intersection Traffic Prediction Using Decision Tree Models," *Symmetry,* vol. 10, no. 9, p. 386, 2018.

[9] N. Polson and V. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies,* vol. 79, pp. 1-17, 2017.

[10] A. E. Retallack and B. Ostendorf, "Current Understanding of the Effects of Congestion on Traffic Accidents," *International Journal of Environmental Research of Public Health,* Vols. 16, 18 3400, 2019.

[11] J. Zeng, Y. Qian, B. Wang, T. Wang and X. Wei, "The Impact of Traffic Crashes on Urban Network Traffic Flow," *Sustainability,* vol. 11, no. 14, pp. 1-14, 2019.

[12] Y. Chung and W. Recker, "A Methodological Approach for Estimating Temporal and Spatial Extent of Delays Caused by Freeway Accidents," *IEEE Transactions on Intelligent Transportation Systems,* vol. 13, no. 3, pp. 1454-1461, 2012.

[13] M. Hashemi and H. A. Karimi, "Weighted machine learning for spatial-temporal data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 13, p. 3066–3082, 2020.

[14] M. Hashemi, "Automatic inference of road and pedestrian networks from spatial-temporal trajectories," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 12, p. 4604–4620, 2019.

[15] M. Hashemi, "Dynamic, stream-balancing, turn-minimizing, accessible wayfinding for emergency evacuation of people who use a wheelchair," *Fire Technology,* vol. 54, no. 5, p. 1195–1217, 2018.

[16] M. Hashemi, "Reusability of the output of map-matching algorithms across space and time through machine learning," *IEEE Transactions on Intelligent Transportation Systems,* vol. 18, no. 11, p. 3017–3026, 2017.

[17] M. Hashemi, "A testbed for evaluating network construction algorithms from GPS traces," *Computers, Environment and Urban Systems,* vol. 66, p. 96–109, 2017.

[18] M. Hashemi, "Intelligent GPS trace management for human mobility pattern detection," *Cogent Engineering,* vol. 4, no. 1390813, 2017.

[19] M. Hashemi and H. A. Karimi, "Collaborative personalized multi-criteria wayfinding for wheelchair users in outdoors," *Transactions in GIS,* vol. 21, no. 4, p. 782–795, 2017.

[20] M. Hashemi and H. A. Karimi, "A machine learning approach to improve the accuracy of GPS-based map-matching algorithms," in *IEEE 17th International Conference on Information Reuse and Integration*, Pittsburgh, PA, USA., 2016.

[21] M. Karlaftis and E. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transportation Research Part C: Emerging Technologies,* vol. 19, no. 3, pp. 387-399, 2011.

[22] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident; analysis and prevention,* vol. 108, pp. 27-36, 2017.

[23] H. Nguyen, C. Cai and F. Chen, "Automatic classification of traffic incident's severity using machine learning approaches," *IET Intelligent Transport Systems,* vol. 11, no. 10, pp. 615-623, 2017.

[24] B. Pan, U. Demiryurek, C. Shahabi and C. Gupta, "Forecasting Spatiotemporal Impact of Traffic Incidents on Road Networks," *2013 IEEE International Conference on Data Mining (ICDM),* pp. 587-596, 2013.

[25] . M. Nowakowska, "Logistic Models in Crash Severity Classification Based on Road Characteristics," *TRB: Journal of Transportation Research Board,* vol. 2148, no. 1, pp. 16-26, 2010.

[26] S. Moosavi, M. H. Samavatian, . S. Parthasarathy and R. Ramnath, "A Countrywide Traffic Accident Dataset," p. arXiv preprint arXiv:1906.05409, 2019.

[27] S. Moosavi, . M. H. Samavatian, S. Parthasarathy, R. Teodorescu and R. Ramnath, "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights," in *International Conference on Advances in Geographic Information Systems*, 27th ACM SIGSPATIAL, 2019.

[28] F. Malin, I. Norros and S. Innamma, "Accident risk of road and weather conditions on different road types," *Accident Analysis & Prevention,* vol. 122, pp. 181-188, 2019.

[29] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," IMS 1999 Reitz lecture, Stanford University, 1999.

[30] S. Yang, J. Wu, Y. Du, Y. He and X. Chen, "Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine," *Journal of Sensors,* vol. 2017, p. 15, 2017.

[31] A. Liaw and M. Wiener, "Classification and Regression by Random Forest," *R-News, ISSN1609-3631,* vol. 2/3, 2002.

[32] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5-32, 2001.

[33] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[34] L. Kang, G. Hu, H. Huang, W. Lu and . L. Liu, "Urban Traffic Travel Time Short-Term Prediction Model Based on Spatio-Temporal Feature Extraction," *Journal of Advanced Transportation,* vol. 2020, 2020.

[35] J. Brownlee, "https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/," [Online].

[36] D. H. Wolpert, "Stacked generalization," *Neural Networks,* vol. 5, no. 2, 1992.

[37] J. Tang, S. Aleyani and H. Liu, Data Classification: Algorithms and Applications, 2014.

[38] "VDOT Traffic Volume," [Online]. Available: https://www.virginiaroads.org/datasets/vdot-traffic-volume.

[39] V. Hurdle, E. Hauser and P. Fargier, "Effects of the choice of departure time on road traffic congestion. Theoretical approach," *Transportation and traffic theory,* vol. 8, pp. 223-263, 1983.

[40] P. J. Hewson, Multivariate Statistics with R, 2009.

[41] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction.

[42] R. Zhou, "How You Can Avoid Car Accident in 2020," https://medium.com.