

Progress Report for STIC Incentive Projects

State: North Carolina

Date: May 13, 2025

STIC Incentive Project Name: Use of AI to Improve Vulnerable Road User Safety

Federal Project Number	Federal Project Fiscal Year	Federal Funds Allocated	Federal Funds Obligated	Federal Funds Expended
	2025	\$125,000	\$125,000	\$0

Name and Title of the State Champion: Daniel Carter, Traffic Safety Specialist

Progress Report Dates (until April 2025): 11/1/24 – 4/30/25

Status (Current progress that has been made on the STIC Incentive Project, including any benefits that have been realized):

Project overview

NCDOT has extensive amounts of data that could be used to improve pedestrian safety. The department collects turning movement counts at over a thousand locations per year around the state. Many of these counts contain the count of pedestrians and bicyclists crossing the road. However, this data is contained in static PDF documents meant only for visual reading by NCDOT staff. As such, many years' worth of data on non-motorist travel exposure are not accessible for analysis on a statewide basis. NCDOT also has crash report data for each non-motorist crash, but the coded values on the crash report do not capture sufficient detail to provide insights into the factors and causes behind the crashes. The reports do contain written text narratives that capture additional data that could be used for further insights. The purpose of this project is to develop a methodology to use artificial intelligence (AI) to assemble data for vulnerable road user safety from these existing DOT data sources.

Progress to date

Turn Count Extraction

- The contractor met with NCDOT to understand the needs for turn count extraction and to walk through the various turn count report formats to understand the variety of ways in which the data is presented.
- NCDOT transferred all historical turn count report files (PDF format) to the contractor, with an approximate total of over 2000 reports.
- The contractor developed an AI model that has been trained to extract Turn Count details from 9 different formats and extracts 18 data points from the Turn Count PDF files, such as location description, weather condition, and duration of count, in addition to the necessary data regarding the number of pedestrians observed in the count.

- All available turn count data was fed into the model, and a summary output dataset has been generated that consolidates the data from over 1,400 turn count reports, encompassing data from all vendors over the past five years.
- The contractor developed a web-based tool to allow NCDOT to upload turn count reports and have the data extracted.
- NCDOT has begun the assessment of the accuracy of the turn count data extraction and the usability of the user interface of the tool.

Crash Narrative Extraction

- The contractor met with NCDOT to understand the needs for the extraction of data from the crash narratives. A series of regular meetings was undertaken to ensure that the contractor understood NCDOT's needs and to refine the data to be extracted, such as the coding of individual data elements to be consistent with current NCDOT data.
- NCDOT set up a Sharepoint site to house the crash narrative data and provided access to the contractor. Via this site, NCDOT transferred data on approximately 28,000 crashes that involved pedestrians. The dataset included crash narrative as well as many other relevant data fields at the crash level as well as unit and person level. NCDOT also provided the dataset of pedestrian crash specific coding that had been conducted post-crash.
- The contractor developed a model to extract four data elements from crash narratives:
 - Non motorist type – this field contained the type of non-motorist involved in the crash, either pedestrian, scooter, or skateboard user.
 - Crash type – this field indicated the specific type of crash, such as dart-out, vehicle failed to yield, and walking along roadway. The coding used in this field was configured to match the coding used in NCDOT's current non-motorist crash data, which uses crash types as laid out in the Pedestrian and Bicycle Crash Analysis Tool (PBCAT).
 - Pedestrian activity – this field was coded in an open ended manner and served to extract a short phrase to describe the pedestrian's activity prior to the crash, such as "ran out into the street", "pushing vehicle", or "standing in the middle of the road".
 - Crosswalk presence – this field indicated the presence of a crosswalk at the crash location, if such information could be derived from the narrative.
- The contractor developed a web interface to allow a user to submit either individual narratives or multiple narratives. If submitted multiple narratives, the interface features bulk extraction of Crash Narratives, processing an input spreadsheet from SharePoint and generating an output file containing the four coded data fields for each crash narrative.
- NCDOT has begun the assessment of the accuracy of the crash narrative data extraction and the usability of the user interface of the tool.