APPENDIX

A. Word importance for severity classification

To estimate word importance in the Incident Description feature, word count matrix has been transformed to a normalized TF-IDF representation (term frequency–inverse document frequency) [29]. N-gram value range is (1,2). Then linear dimensionality reduction has been performed using truncated singular value decomposition to 50 componenets for 7 iterations. Then we used GBDT classification model to fit incident severity and three quantiled groups (ratio 33%:33%:33% to represent equaly sized groups with duration intervals 0-29min, 30-71min and 72-2750min) of the incident duration. Classifer predictions were then analyzed for feature importance using LIME method [30], where every feature represents 1 word or 2 word combination presence in the incident description. One or more combinations of word in the description can contribute to the incident being classifed into one of severity groups (Fig. 9) - presence of "lanes blocked" and "two lanes blocked" has the highest contribution to the incident being classifed into highest (3) or lowest (0) severity group. Severity 1 or 2 is more related to the actual location, which represented as word describing Cesar Chavez St and I-280 Interstate Highway. High positive and opposite high negative contribution of words towards severity group observed for severity groups 1 and 2, where "280" and "chavez" have high opposite contributions, making this groups easily separable. When we perform classification towards equaly sized incident duration groups, "lanes blocked" has the highest positive contribution of the incident to be classified into low duration group. If accident happens on Cesar Chavez St, it can be easily classified into low duration group signifying importance of location for the task of incident duration prediction. High negative contribution of "lanes blocked" observed for duration group 1 with the highest contribution of "280" word meaning that incident appears on I-280 Interstate Highway.

B. Traffic flow and traffic speed on the day of the incident

The following plots represent recorded traffic speed and flow on the day of the incident and week before in 500m proximity of the incident along the road (see Fig. 11 and 12). Reports in CTADS data set indicate that the highest impact of traffic incident is attributed to significant decrease in traffic speed, while traffic flow stays the least affected by disruption.

Severity group=0		Severity group=1		Severity group=2		Severity group=3	
Weight?	Feature	Weight?	Feature	Weight?	Feature	Weight?	Feature
+0.644	lanes blocked	+1.559	chavez	+2.805	280	+0.982	lanes blocked
+0.345	two lanes blocked	+1.190	cesar	+1.909	280 northbound	+0.427	two lanes blocked
+0.231	due	+0.973	<BIAS>	+0.828	blocked	+0.365	due
+0.034	due to	+0.894	st	+0.740	to accident	+0.174	on i
+0.007	to accident	+0.475	i	+0.736	accident	+0.110	due to
-0.407	lanes	+0.467	to	+0.721	i 280	+0.008	to accident
-0.620	blocked	+0.465	on	+0.697	chavez st	-0.076	cesar chavez
-0.689	i	+0.351	at	+0.677	accident on	-0.127	lanes
-0.704	<BIAS>	+0.309	northbound	+0.448	two lanes	-0.546	blocked
-0.748	to	+0.307	cesar chavez	+0.375	lanes blocked	-0.621	i
-0.760	st	+0.289	two	+0.336	at cesar	-0.666	on
-0.769	accident	+0.153	due	+0.194	due to	-0.672	<BIAS>
-0.793	two	-0.031	northbound at	+0.187	blocked due	-0.678	to
-0.797	due	-0.101	blocked due	+0.138	lanes	-0.710	at
-0.800	on	-0.125	due to	+0.070	northbound at	-0.762	two
-0.818	at	-0.310	at cesar	+0.022	on i	-0.773	due
-0.869	northbound	-0.330	two lanes	-0.160	northbound	-0.918	accident
-0.924	280	-0.372	lanes	-0.208	due	-0.953	st
-0.979	chavez	-0.466	lanes blocked	-0.354	cesar chavez	-0.961	cesar
-1.149	cesar	-0.647	accident on	-0.358	at	-0.997	chavez
		-0.684	i 280	-0.369	two	-1.116	280
		-0.692	chavez st	-0.498	on	-1.140	northbound
		-0.711	accident	-0.509	to		
		-0.728	to accident	-0.534	i		
		-0.913	blocked	-0.891	st		
		-1.993	280 northbound	-0.994	<BIAS>		
		-2.728	280	-1.110	cesar		
				-1.479	chavez		

Fig. 9: Word importance estimation using LIME method for incident severity groups

duration group=0		duration group=1		duration group=2	
Weight?	Feature	Weight?	Feature	Weight?	Feature
+1.307	lanes blocked	+0.548	280	+0.389	chavez st
+0.653	two lanes	+0.444	northbound	+0.256	280 northbound
+0.461	blocked due	+0.357	blocked	+0.149	blocked due
+0.422	lanes	+0.218	chavez	+0.132	northbound at
+0.326	to accident	+0.214	st	+0.092	at cesar
+0.324	on i	+0.213	accident	+0.075	cesar chavez
+0.255	at cesar	+0.182	cesar chavez	+0.068	to accident
+0.230	due to	+0.095	two lanes	+0.062	cesar
+0.216	northbound at	+0.091	cesar	+0.017	to
+0.211	chavez st	+0.050	due to	-0.036	<BIAS>
+0.177	accident on	+0.039	i 280	-0.057	lanes blocked
+0.026	i 280	+0.034	lanes	-0.080	due
-0.123	st	+0.029	280 northbound	-0.088	at
-0.153	cesar chavez	-0.013	on	-0.133	two lanes
-0.232	blocked	-0.030	<BIAS>	-0.232	accident
-0.232	280 northbound	-0.037	two	-0.264	chavez
-0.275	i	-0.069	to accident	-0.383	st
-0.290	at	-0.072	northbound at	-0.502	northbound
-0.348	on	-0.077	i	-0.580	lanes
-0.405	chavez	-0.129	blocked due	-0.594	280
-0.437	northbound	-0.204	chavez st	-0.633	blocked
-0.439	280	-0.655	lanes blocked		
-0.440	to				
-0.449	due				
-0.485	accident				
-0.544	two				
-0.724	cesar				
-0.918	<BIAS>				

Fig. 10: Word importance estimation using LIME method for incident duration groups

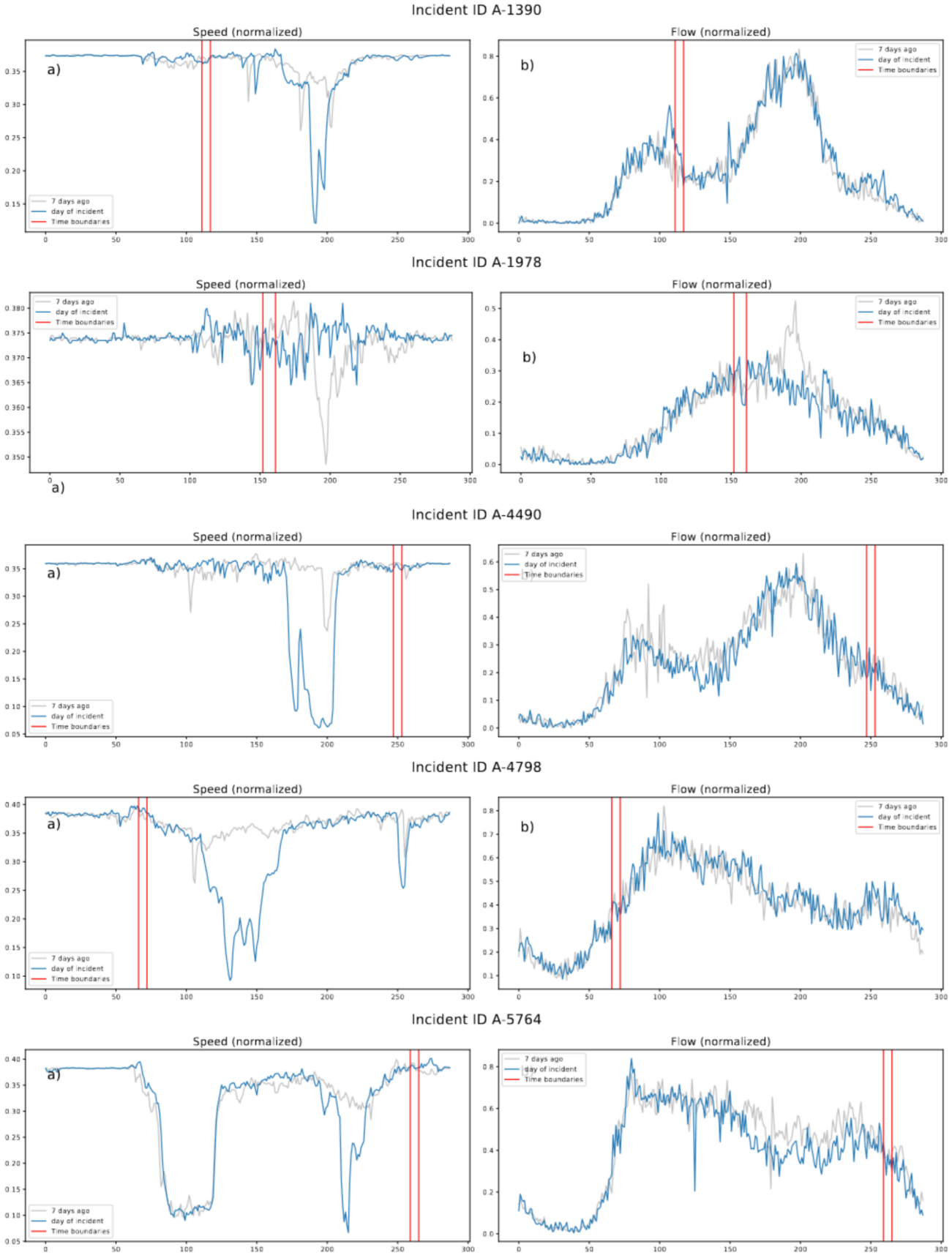

Fig. 11: Traffic speed and flow during the day of the incident. Part #1

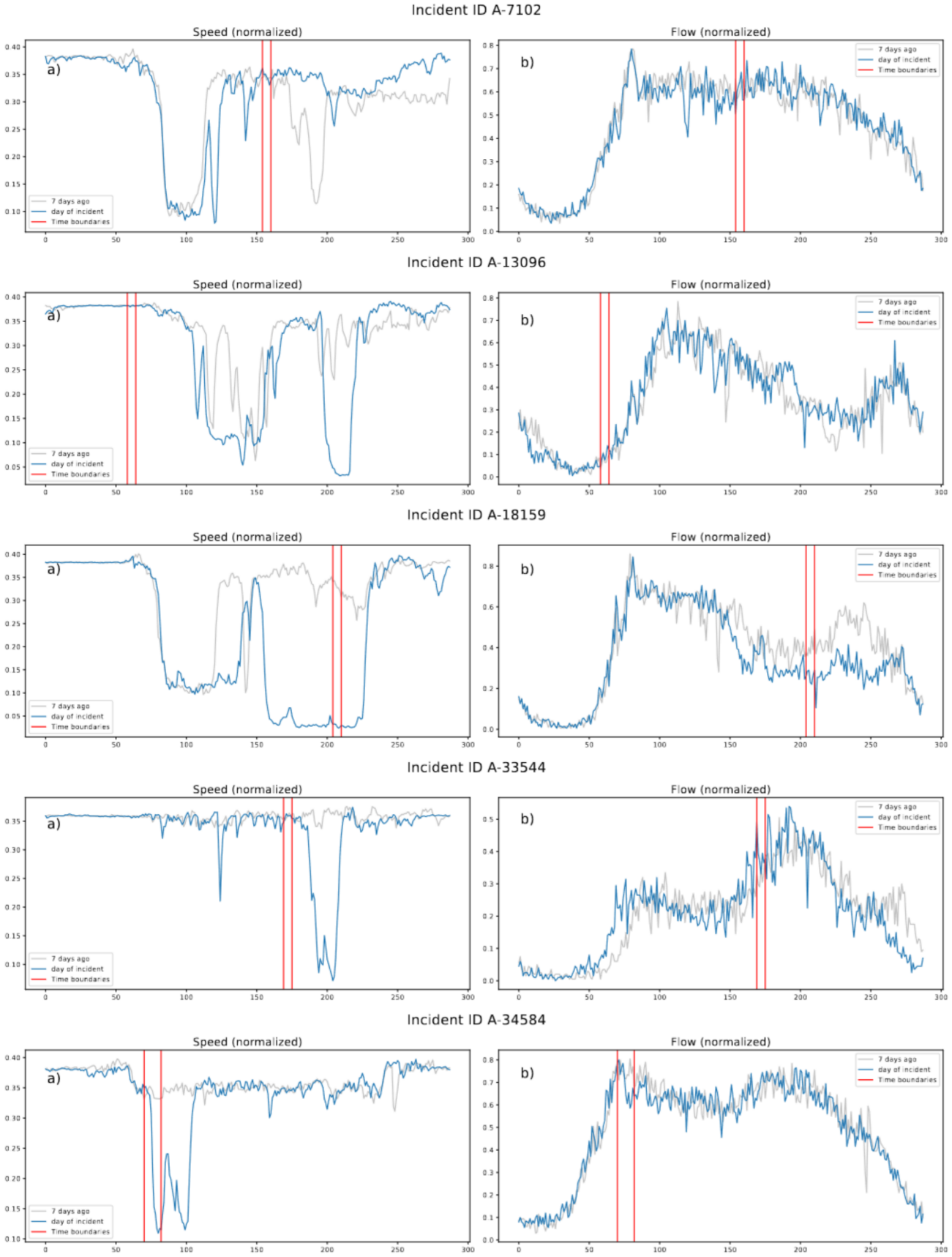

Fig. 12: Traffic speed and flow during the day of the incident. Part #2