**INTRODUCTION**

Trains as public transportation have been considered as safer than other means. However, passengers on trains stations sometimes face many risks because of many overlapping factors such as station operation, design, and passenger behaviors. Due to the gradually increasing demand and the heavily congested society and the state of some station’s layout and complexity in design, there are potential risks during the operation of the stations. Furthermore, Passenger, people and public safety is the main concern of the railway industry and one of the critical parts of the system. European Union put into practice Reliability, Availability, Maintainability and Safety (RAMS)as a standard in 1999 known as EN 50126. Aiming to prevent railway accidents and ensure a high level of safety in railway operations. The RAMS analyses concepts lead to minimizing the risks to acceptable levels and rise safety levels. However, that have been an urgent issue and still, the reports show several people are killed every year in the railway station, some accidents lead to injuries or fatalities. For example, In Japan in 2016, 420 accidents occurred that included being struck by a train, which resulted in 202 deaths. This including of those 420 accidents, 179 (resulting in 24 fatalities) included falling from a platform and following injury or death as a consequence of hitting with a train [1]. In the UK, 2019/20, it has been reported that Most passenger injuries occur from accidents in stations. Greatest Major injuries are the outcome of slips, trips and falls, of which there were approximately 200 [2] play significant impact in reducing injuries on station platforms and provide quality, reliable and safe travel environment for all passengers, worker and public. Even if some accident does not result in deaths or injuries, such accidents cause delay, cost, fear and anxiety among the people, interruption in the operations and damage the industry reputation. Also, to provide or invest any control safety measurements the stations it is crucial to considering the risks associated with the railway incidents and risks in the station and identification of many factors related to the accident by a comprehensive knowledge of the root cause of accidents considering all the possible technology.

The objective of this research is to analysis a collection case of accidents between 01/01/2000 and 17/04/2020 data to introduce a smart method, which expected to develop the safety level future, the risk management process, and the way to collect data in the railway stations. This data been gathered by RSSBS and agreed to be used for the research purpose. Analyzing an extensive amount of data recorded in a different form are a challenging job. Nowadays, it is hard to obtain for specific information in such mix digitization big data in including Web, video, images and other sources, it is research of a needle in a haystack. Thus, a powerful tool for assistance manage, search and understand these vast amounts of information is needed indeed [3], [4]. Many pre-processing techniques and algorithms are required to obtain valuable characteristics from an enormous amount of safety data in the stations including textual. The study covers the topic modeling to identify useful characteristics such the root cause of the accidents and also exploring the factors which are multiple groups of words or phrases that explain and summarize the content covered by an accident’s reports reducing time with high accuracy of outcomes. Topic modeling techniques are robust smart methods that extensively applied in natural language processing to topic detection and semantic mining from unstructured documents. Consequently, It has been suggested in this work the LDA model which is one of the best-known probabilistic unsupervised learning methods that marks the topics implicit in collection of contexts [5]. Since increasing of applying new technologies and the revolution of data, the development of technology and utilizing AI in many fields it suggested in this paper a smart analysis utilizing the topic modeling techniques which can be very useful and effective to semantic mining and latent discovery context documents and datasets. The other source of data (Images-videos and numerical) been conducted utilizing AI approaches which cover supervised learning [6], [7], so the unstructured textual data is targeted.

Hence, our motivation is to investigate the topic modeling approaches to risks and safety accident subjects in the stations. This work provides the method of topic modelingbased on LDA with other models for advanced analytics, aiming to make contributions in the future of smart safety and risk management in the stations. Through applying the models, we investigate the safety accidents for fatality accident in the railway.

This paper establishes an innovative method in the area to studies how the textual source of data of railway station accident reports could be efficiently used to extract the root causes of accidents and establish an analysis between thetextual and the possible cause. Where the full automated process that has ability to get the input of text and provide outputs not yet ready [8]. Applying this method expected to come overcome issues such as aid the decision-maker in real time and extract the key information to be understandable from non-experts, better identify the details of the accident in-depth, design expert smart safety system and effective usage of the safety history records. A Such results could support in the analysis of safety and risk management to be systematic and smarter. Our approach uses state-of-the-art LDA algorithm to capture the critical texts information of accidents and their causes. The rest of this paper is arranged as follows: In Section II, related work in both accident analysis and text classification with deep learning have been presented. Section III describes in detail the approach that has been used along with evaluation criteria. Section IV provides details of our implementations and section V reports the results. Finally, Section VI presents the conclusion.