

Abstract No: PO-26

Identification of factors leading to elephant deaths in human-elephant conflicts

W. A. D. M. Lakshitha^{1*}, N. V. Chandrasekara¹, H. W. B. Kavinga¹, N. Withanage²

¹ Department of Statistics & Computer Science, University of Kelaniya, Sri Lanka

² Department of Statistics, University of Sri Jayewardenepura, Sri Lanka
lakshitha3536@gmail.com *

Human-elephant conflicts (HEC) have emerged as one of the main challenges that Sri Lanka faces throughout several decades. According to the official data of the Department of Wildlife Conservation (DWC), the number of elephant deaths is higher than the number of human deaths due to HEC per year. This research focused on the North Central Province, where the highest number of elephant deaths have been recorded. Hence, the objectives of this research are to identify the main factors that have affected the deaths of elephants and to identify suitable models to predict the causes of elephant deaths due to human-elephant conflict. Although there has been much research related to HEC worldwide, no published research studies were found in the literature that utilized advanced statistical techniques such as Multinomial Logistic Regression (MLR), LASSO regression, Decision Tree (DT), Support Vector Machine (SVM), and Probabilistic Neural Network (PNN) for their studies. However, this research will address that research gap by constructing models for classifying the causes of elephant deaths resulting from HEC. Data was collected from various departments, including DWC, the Department of Meteorology, and the crop calendar of the Department of Agriculture. Furthermore, Pearson's Chi-square and Fisher's exact tests were used to identify the association between the cause of death and influencing factors. Five variables, including the elephant age group, grass levels, gender, rainfall season, and place of death, were found to significantly influence the causes of death of an elephant. MLR and Data Mining (DM) techniques were initially utilized, but due to multicollinearity arising in MLR, the LASSO technique was employed as a remedial method. To overcome the class imbalanced problem, 90% of the data were randomly selected for model building while maintaining the class ratio of the response variable, and the remaining 10% of the data were used for testing. Performance measures, overall classification accuracy (OCA), and Misclassification Percentage of Critical Cases (MPCC) were used to evaluate and compare the classification potential of models. Models such as final MLR, LASSO, DT, SVM with Polynomial and Gaussian Kernels, and PNN with spread 0.801 illustrated 42.30%, 50%, 53.84%, 69.23%, 73.07%, and 73.07% of OCA. In addition, the above models showed 34.61%, 30.76%, 7.69%, 11.53%, 19.23%, and 26.92% MPCC respectively. Finally, the SVM model with Gaussian Kernel exhibited high OCA (73.07%) with 19.23% of MPCC as the better model since the PNN showed a high MPCC of about 26.92%. These findings will be helpful for authorities in their future and existing projects.

Keywords: Causes of Elephant Death, Multinomial Logistic Regression, LASSO Regression, Data Mining Techniques