During the last few decades, many efforts have been made to assess the reliability of nondestructive evaluation (NDE) technologies used for the detection of subsurface damage in concrete bridge decks. During these efforts, reliability of NDE technologies has frequently been relegated to anecdotal discussion, the probability of detection (POD) estimation, or accuracy assessment. Although these indices are important, most of the previous work did not take account of the probability of false alarm (POFA) of NDE technologies, nor did they investigate the reliability considering different threshold settings to classify NDE results to a dichotomous scale. Consequently, the assessments are rather controversial, and there is no general agreement about the reliability of such technologies. Because most NDE technologies are characterized by noisy data and less than-perfect detection characteristics, reliability is to be carefully assessed considering all practical threshold settings and accounting for all possible diagnosis output. In other words, when NDE data do not fall into either of the two obviously defined categories: true positive (TP), meaning the NDE data indicates a defect and there is a defect, or true negative (TN), meaning the NDT data indicates no defect and there is no defect, reliability analysis should also include the two types of incorrect indications: failure to give a positive indication in the presence of a defect (false negative, FN) and giving a positive indication when there is no defect (a false alarm or false positive, FP).

In this research, the reliability of impact echo (IE), infrared thermography (IRT) and ground penetration radar (GPR) technologies is assessed by using a statistical analysis method called receiver operating characteristic (ROC). These three NDE technologies are the most widely technologies used for the detection of concrete subsurface delamination. The proposed analysis method has the capability to integrate POD and POFA indices over a wide range of decision threshold settings into a single curve which is useful in assessing the trade-off in choosing a threshold and for quantitatively comparing the performance of NDE technologies. This methodology for assessing NDE reliability is intended to provide a more effective means of comparing different technologies used in civil engineering applications, to make the evaluation of quantitative scheme, to reduce subjectivity and variability in interpreting NDE data, and to improve sensitivity to extract more information from NDE data. Area under ROC curve (AUC) which is interpreted as the probability of correctly classifying an arbitrarily pair of negative and positive test points can provide for the desired quantitative reliability index which can be used to quantitatively compare the performance of one NDE technology to another.

Results of this research obtained from ROC analysis describe the capability of IE and IR in detecting subsurface fracture damage such as delamination and debonding. In both technologies, there exist some threshold settings that can provide for a relatively high POD with very low POFA, and consequently, the areas under their ROC curves were very high. Data obtained from GPR testing, in contrast, indicates that GPR technology has a very limited ability to detect physical damage such as subsurface delamination and its ability is limited to only the detection of corrosive environment such as moisture and chloride when the concentrations of these factors are above some threshold values that may facilitate the initiation of steel reinforcement corrosion.