**Chapter 4: Validation**

* 1. **Introduction**

Our study addresses three main research questions, which are defined below. We also explain how our experiments are designed to address them. The goal of the study is to evaluate the efficiency of our approach for generating correct detection rules while minimizing the rules-complexity and maximizing the number of detected interface defects.

* 1. **Research Questions**

The three research questions are:

**RQ1**: To what extent can the proposed approach enhance the graphical user interface quality by the detection of defects?

**RQ2**: To what extent can the adaptation of multiple metrics impact on better investigation for defects in the system?

**RQ3**: To what extent can the proposed approach minimize the number of needed detection rules?

**RQ4**: How does the proposed NSGA-II-based algorithm perform compared to a mono-objective approach?

To answer RQ1, we (…).

To answer RQ2, we (…).

To answer RQ3, we (…).

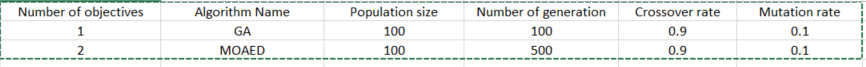
To answer RQ4, we (…).

* 1. **Studied Project**

The validation is being conducted over the evolution of different open source Java GUI systems [reference][ reference][ reference][ reference][ reference]. The corpus used includes releases of (…). Table X gives an overview about the studied systems. The table also reports the number of code smells identified manually in the different systems. This number will be later used to calculate the precision and recall of the given approach. We have chosen these projects because of their small to medium size and because they contain multiple GUIs that can be used as input to our approach.

* 1. **Experimental Setting**

Since search algorithms are stochastic, they may generate various results for the same problem instance in multiple simulations. That’s why, our experimental study is performed throughout 31 independent runs for each problem instance and the obtained results are statistically analyzed by using the Wilcoxon rank sum test [reference] with a 99% confidence level (α = 1%). This guarantees that the obtained results of two algorithms used in this experiment are samples from continuous distributions with equal medians. For each input system, we compute the p-value obtained by comparing NSGA-II and mono-objective search results to determine whether the performance difference between them is statistically significant or just a random result. The following Table Y shows the tuning configuration for the two algorithms used in our experiments



* 1. **Research Questions**
     1. **Results for Research Question 1**

Insert the suitable excel tables here

* + 1. **Results for Research Question 2**

Insert the suitable excel tables here

* + 1. **Results for Research Question 3**

Insert the suitable excel tables here

* + 1. **Results for Research Question 4**

Insert the suitable excel tables here

* 1. **Conclusion**

In this chapter, we present our two contributions of our research which are the proposed metrics of adaptive user interface evaluation and the meta-heuristic used to generate a set of evaluation rules that evaluate a several adaptive interfaces.

In the next chapter, we will discuss the experimentation of our evolutionary algorithm. Moreover, we will present a comparison study of our used multi-objective evolutionary algorithm with another mono-objective evolutionary algorithm.