Rough set theory has been developed as a means to analyse vague description of objects. Objects characterized by attributes may be indiscernible based on the information available about them. Rough sets are approximation representations of a given set in the form of lower and upper approximations derived from crisp partitions. Rough set theory approach is important in the areas of machine learning, knowledge acquisition, decision analysis, and knowledge discovery from databases.

The original rough set approach is restricted to the case where objects attributes in information systems are described by precise values. Actual applications, however, often contain imperfect data including but not limited to missing, uncertain and imprecise values.

Though numerous approach dealing with missing values have been published in the literature, a lack of solutions to solve issues of uncertainty and imprecision still remains. This research aims at proposing possible solutions for all of imperfect data mentioned above. The work contains introducing a representation of imperfect data, proposing two new rough set models and discussing methods for acquiring knowledge in imperfect information systems.

First of all, the research studies imperfect information systems. There are many reasons why data is missing. Some attribute values are not recoded because they might not be necessary. Information also does not exist, although it is important. This is because of data may not have been collected or it may have been deleted accidentally. In incomplete information systems that have missing values, some approaches published in the literature are reviewed. These models have problems that may lead to poor approximations or lost information.

Imprecision, on the other hand, provides some information about the real data by a set of possible values. Specific kinds of imprecise information include disjunctive information, negative information, range information, and error margins.

The last considered type of imperfect data is uncertainty. A value is uncertain when it is denoted by a value and its probability. Lost values with pre-defined probability distribution are also considered as uncertainty.

To propose new rough set models in imperfect information systems, a representation of imperfect values is introduced. This representation must have ability to present any type of imperfect data. A solution chosen in this research is a combination of transforming missing, uncertain and imprecise values to probabilistic data.
The second work of the research is using the representation of imperfect data to define a rough set model based on valued tolerance relations. The skeleton of a probability based solution consists of determining probability distribution of values in domains, calculating probability of matching between two objects, and obtaining a tolerance/similarity degree.

For this kind of relations, this research first suggests two methods for estimating probabilities of object attribute values based on frequencies of values appearing on data sets. It then illustrates methods to obtain probabilities of matching - the probability that two objects are tolerant of each other on an attribute - for imperfect data. Combining these probabilities with another index called equivalence existence and using thresholds for controlling uncertainty levels, a valued tolerance relation is proposed. This relation can avoid problems stated in the literature of several rough set models.

The third work is proposing a rough set model based on Dempster-Shafer theory. Taking advantages of Dempster-Shafer theory that allow us to assign a probability mass to a set of events into account, several basic relations that are determined by comparing possible values sets of two objects on an attribute are first defined. Mass assignments for the occurrences of those basic relations on an attribute are also calculated from probabilities of possible value sets. Considering each attribute as a source of evidence and employing combination rules, mass assignments of basic relations for a set of attributes are obtained. Calculating and comparing belief and plausibility measures of hypotheses on basic relations, equivalence, tolerance and similarity relations among objects are then determined. The approach based on Dempster-Shafer theory is useful when there is no information about probability distribution of attribute values.

The last work is introducing methods of knowledge discovery in imperfect information systems. Usually, a discernibility matrix is used to calculated reducts and core of an information system. To induce decision rules, an algorithm named LEM2 is a famous solution. However, it is evident that those approaches cannot be used in some cases. Therefore, finally, methods to obtain reducts and core and to induce decision rules in imperfect information systems are discussed. In this discussion, the algorithm of deriving rules is able to obtain certain and possible rules by calculating lower and upper covering of possible rule sets.

The study also points out some limitations and addresses three directions of further works.