CHAPTER 6
CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION:

Since Codd introduced the relational model in the early 1970s, many attempts have been made to incorporate imperfect information (including incomplete, imprecise and uncertain information) in relational databases. However, the results obtained after three decades of the research are not really significant. Representing and manipulating this kind of imperfect information in a semantically correct way has raised theoretical problems of enormous complexity. With the advances in artificial intelligence, logic programming, expert systems, deductive databases, the need for manipulating imperfect information has even been furthered. There have been multiple approaches in each of these areas but general solutions are not known.

The work reported in this thesis surveys different sources of imperfect information in relational databases and application of artificial intelligence. It also highlights the usefulness of imperfect information as a basis for efficient and effective decision-making. The objective was to analyze uncertainty with the essential purpose of obtaining practical results. In this respect, a medical expert system was developed that uses a relational database as its knowledge base and applies probabilistic analysis instead of if-then-else rules in vogue. As per the symptoms communicated by the patient, the expert system will respond with the possible diseases the patient is suffering from in the increasing order of their probabilities.

The architecture of a conventional database management system has been extended so as to represent and manipulate uncertain information using the theory of probability. A comprehensive probabilistic relational database management system (PRDBMS) has been proposed that is capable of representing and manipulating uncertain information by using the theory of probability. Since, the structure of any information system without corresponding operator or inference techniques is just like anatomy without physiology,
therefore various theoretical issues related to developing an exact prototype of the proposed system have been addressed. The probabilistic integrity constraints, probabilistic relational algebra, probabilistic normal forms were also defined likewise. The query language for the comprehensive probabilistic relational database management system named Probabilistic Structured Query Language (PSQL) has been presented and its syntax has been provided using BNF notation.

There were two reasons that led to the study of uncertainty in the context of relational databases as indicated below:

1. Owing to the complexity of the problem and in order to fulfil the objectives, only a particular case was chosen without any pretension to find the general solution applicable to every discipline.

2. Owing to richer theoretical background of relational model, the objective was to study uncertainty with the essential purpose of obtaining practical results. The objective was to be able to translate the conceptual analysis in terms of methods or algorithms.

Since a database system must ensure proper execution of transactions, the similar framework has been defined for the comprehensive probabilistic relational database management system. As the probabilistic database is maintained in the form of probabilistic relations, a probabilistic transaction has been defined likewise that shall be performing various operations on the probabilistic database. The PLACID properties that the transaction should possess to maintain integrity of data have been identified and enumerated. The acronym PLACID stands for Probability Preserving, Lossless, Atomicity, Consistency, Isolation and Durability characteristics that should be exhibited by the transaction. To maintain consistency among the concurrent transactions, an algorithm has been provided that is based upon locking technique for handling various issues related to concurrent execution of the transactions.

To avail of the benefits associated with distributed processing and to respond to various real-life situations, the probabilistic database has been designed to be distributed across a
network and various issues of query execution and distributed concurrency control have been addressed for such an environment.

6.2 FUTURE SCOPE:

As a step further towards the research carried out in the thesis, attempts can be made to incorporate uncertainty at conceptual level like evolving probabilistic ER (Entity-Relationship) model. As far as concurrency control techniques are concerned, the same analysis needs to be carried out for timestamp based ordering algorithms. Certain optimistic concurrency control algorithms also need to be considered for probabilistic databases.

Deadlock management with respect to probabilistic databases is an open problem. Algorithms need to be developed for deadlock prevention, deadlock avoidance, deadlock detection and resolution. Some of the aspects have been studied by distributing the probabilistic relations over a network. As a further step, the reliability issues need to be addressed to a greater depth. The protocols need to be developed that will ensure reliability of the distributed probabilistic databases. It would be interesting to study the complexity of query reliability by analysing various algorithms on the basis of their time and space complexity. Query optimization with respect to probabilistic databases is an open problem.

All these issues need to be addressed to evolve a practical version of the prototype probabilistic relational database management system.