DECLARATION

I, Mr. Bhavani Sankar Panda, a scholar in the department of Computer Science and Engineering, Centurion University of Technology and Management, Paralakhemundi, Odisha, hereby declare that the thesis entitled "**Expert System Development for Retrieval of Missing Data through Soft Computing Techniques**" being submitted to the Faculty of Engineering & Technology, Centurion University of Technology and Management, is a record of the my own work, carried out under the supervision of Dr. Sasanko Sekhar Gantayat and Dr. Ashok Misra. I also declare that the matter embodied in this thesis has not been submitted for the award of any other degree either to this university or any other university or institute.

Bhavani Sankar Panda

CERTIFICATE

Certified that the thesis entitled **"Expert System Development for Retrieval of Missing Data through Soft Computing Techniques"** being submitted to the Faculty of Engineering & Technology, Centurion University of Technology and Management, Paralakhemundi, by Mr. Bhavani Sankar Panda in fulfillment of the requirements for the award of the degree of "Doctor of Philosophy" in Computer Science and Engineering is a record of the scholar's own work, carried out by his under our supervision and guidance. We certify that to the best of our knowledge, the matter embodied in this thesis has not been submitted for the award of any other degree either to this university or any other university or institute.

Dr. Sasanko Sekhar Gantayat
Dept. of Computer Science & Engg., Technology,
GMR Institute of Technology,
Rajam, Andhra Pradesh.

Dr. Ashok Misra
Centurian University of

Parlakhemundi, Odisha.

ACKNOWLEDGEMENTS

I am greatly indebted to express my sincere thanks and deep gratitude **The Almighty God** for blessing, protecting and guiding me throughout this period. I could never have accomplished this without the faith I have in the God.

I express my profound sense of reverence to my supervisor **Prof. Dr. Sasanko Sekhar Gantayat**, for his constant guidance, support, motivation and untiring help during the course of my Ph.D. His in-depth knowledge on a broad spectrum of Computational Intelligence topics has been extremely beneficial for me. He has given me enough freedom during my research, and he has always been nice to me. I will always remember his calm and relaxed nature. I am thankful to the Almighty for giving me a mentor like him.

I express my deepest gratitude to my co-supervisor **Prof. Dr. Ashok Misra**, Centurian University of Technology and Management, Parlakhemundi for boosting my morale throughout the course of research. He has always been caring, a source of wisdom and motivation. He is a great leader.

I would also like to thank the members of my Doctoral Committee, **Dr. M. R. Senapati** (CIT, Jatni Campus), **Dr. Anita Patra** (Dean SOET, PKD), **Dr. P.S.V Ramana Rao** (RC Member, CUTM) and **Dr. M.L. Narasimham** (Dean, Academic Research) for accepting my submission of this thesis.

I would also like to thank all the member faculties of my department for their source of help and support, not only during the thesis work, but also throughout my Ph.D. tenure. I would like to thank all those who directly or indirectly helped me throughout the project.

At last I would like to thank my wife **Meenaksh**i and my son **Akash** for their love, affection, help and support without which this project has been an incomplete one.

Bhavani Sankar Panda

CONTENTS

CHAPTER NO. **TITLE** **PAGE**

NO.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

RF Set	Rough Fuzzy Set
MCAR	Missing completely at random
MAR	Missing at random
MNAR	Missing not at random
ML	Maximum Likelihood
MI	Multiple imputations
LERS	Learning from Examples based on Rough Sets
PVI	Predictive Value Imputation
DBI	Distribution-Based Imputation
UVI	Unique-Value Imputation
LEM2	Learning from Examples Module, version 2
MLEM2	Modified Learning from Examples Module, version 2
RST	Rough Set Theory
CBRS	Covering Based Rough Set
SS	Soft Sets
KB	Knowledge Base
AI	Artificial Intelligence

ABSTRACT

Missing data is a familiar and unavoidable problem in large datasets and is widely discussed in the field of data mining and statistics. The impact of missing data on quantitative research can be serious, leading to biased estimates of parameters, loss of information, decreased statistical power, increased standard errors, and weakened generalizability of findings. All researchers have faced the problem of missing quantitative data at some point in their work. Research informants may refuse or forget to answer a survey question, files are lost, or data are not recorded properly. Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data.

Many techniques for handling missing data have been proposed in this literature. Most of these techniques are overly complex. This thesis explores an imputation technique based on rough set computations. In this thesis, characteristic relations are introduced to describe incompletely specified decision tables. It is shown that the basic rough set idea of lower and upper approximations for incompletely specified decision tables may be defined in a variety of different ways. Empirical results obtained using real data are given and they provide a valuable and promising insight to the problem of missing data. Missing data were predicted with an accuracy of up to 99%.

The data base of a Knowledge-Based System (KBS) contains inaccuracies and uncertainties which are inherent in the description of the rules given by the expert. They are due to the difficulty of representing the facts involved in the antecedents and the consequents of the inference rules, which are expressed in most cases by ambiguous characterizations (such as a color or an age described as “young”), or by imprecise data, (for instance, a length described as “approximately equal to 15 meters”). Uncertainties appear, particularly when the expert is not certain of the validity of the rule in any cases. Another problem stems from the

utilization of these rules when the observed facts are not identical with the condition expressed in their premises, but are not too different from them. In a process using classical logic, the rule would not work in this case, but the fact is sometimes so close to the characterization indicated in the premise of a rule that it seems interesting to obtain a deduced fact, even if we must restrict its validity through a coefficient of uncertainty.