Datasets are getting complex day by day. A decade ago, a futuristic approach was introduced to deal with datasets consisting of single valued features, interval valued features, multivalued features, distribution features etc. Modeling such heterogeneous dataset for knowledge extraction was termed as symbolic data analysis (SDA). From then on, many techniques were proposed to handle the above variety of features. A decade later the scenario is still more challenging. The drastic development of database technology towards object relational models has brought in samples/objects characterized by temporal data attributes, image attributes, video attributes etc. Thus the heterogeneity of the dataset has been increasingly intricate, and such a complex dataset is termed as generic dataset according to this thesis.

Most of the knowledge mining models define algorithms to handle a particular type of feature/attribute. When heterogeneous dataset is encountered, similar feature/attribute types are grouped together and proximity between the samples are measured based on the individual group with specific distance measures. Then the distances are summed to provide a comprehensive proximity between the samples. This thesis makes an attempt to explore the possibility of measuring the comprehensive proximity with a generic model which is capable of handling all kinds of features. Algorithms capable of modeling generic datasets with a generic model for knowledge extraction are categorized as Generic Knowledge Mining techniques.

Feature transformation is an important stage of knowledge mining for proposing generic models. In symbolic data analysis, histogram feature was successfully found to characterize the other features types like interval valued, single valued, multivalued etc. Although the proposal of feature transformation to histogram was interesting, it was not very well acknowledged due to the increase in computational complexity in terms of histograms. Thus the exploration towards transformation to generic histogram feature type and also processing of histogram objects became slack. This thesis continues the journey of histogram data type as a generic feature type by proposing dimensionality reduction models through principal component
analysis, supervised and unsupervised learning models, classification and clustering models for histogram dataset, neural network models and subsequently wavelet histogram network models. The thesis also attempts to handle the complexity issue pertaining to histogram objects. A regression line based model to reduce the computational complexity of histogram models is proposed. The foundational ideas of regression line are extended to propose regression line PCA for dimensionality reduction, regression line neural network models for classification and clustering, and in the sequel, wavelet regression line networks for mining media databases.

With object relational database becoming popular day by day the heterogeneity of dataset has further increased through the incorporation of temporal data features, image features, video features etc. These media features are generally complex data structures made of huge samples. Histogram objects have been commonly used for summarizing the characteristics of these huge samples for data analysis. But this summarizing process results in a very high information loss. This is because of the retention of only the coarse information and loss of finer information along the transformation process. The information loss could be minimized if these media data are analyzed at multiple resolutions before summarization into histograms. Wavelet has been a successful tool for multiresolution analysis. Thus the thesis makes vital attempts to transform these object features into histogram features through wavelet analysis. Such wavelet histograms have resulted in high quality knowledge extracts.

The insight into wavelet analysis has also led to the proposal of a new transform named \textit{WaveSim Transform}. The interesting theory behind this novel transform has been used for multiresolution mining of temporal data.

By modeling the media features like temporal data, image data, video data through wavelet/wavesim histograms and adopting the feature transformation process suggested by SDA (Symbolic Data Analysis) group, the thesis transforms the entire heterogeneous dataset into a generic histogram dataset and proposes models to process it. Each of the histogram models has been supplemented with a regression line model for low complexity knowledge mining.