This work is concerned with the use of soft computing algorithm to unravel the distinct features of a class of protein molecules known as knot proteins. In algebraic topology, knots are defined as closed curves in three-dimensional Euclidean space $\mathbb{R}^3$ and are categorized according to the minimal number of crossings in a projection onto a plane. Protein knots are interesting structural motifs. The number of protein submissions to Protein Data Bank (PDB) with knots in their native structures, is increasing. The structural biology community is gradually working its way towards understanding their role in structure and function determination. Enzymatic activity is one of the major activities of a knot in protein. The knotted regions have been shown to be important in ligand binding too. Knotted structures are difficult to fold, but once folded they maintain the conformation of the folded state which contribute to thermal stability.

Knots are examples of topologically nontrivial structures in proteins with potentially very high biomedical relevance. They may even have some substantial biomedical significance in relation to illnesses such as Parkinson's disease. Ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) (EC 3.1.2.15) is a deubiquitinating enzyme. A point mutation (I93M) in this gene encoding this protein is implicated as the cause of Parkinson's disease. Furthermore, a polymorphism (S18Y) in this gene has been found to be associated with a reduced risk for Parkinson's disease. The gene is also associated with the Alzheimer's disease, and required for normal synaptic and cognitive function. Hence further studies in knot proteins may provide relevant insightful methods in drug discovery for such diseases. For proteins, knot theory is likely to be useful, because the geometry and motions of protein backbones may be modeled using techniques from knot theory. Two proteins or subsegments of proteins are similar if there is a motion that transforms one into the other.
while avoiding backbone self-collisions. Knot invariants help to assess the similarity of proteins. This invites the need to develop study about knot invariants and it could be taken up as a future work since they play an important role in finding similarity of proteins. The present work builds upon several special parameters associated with knot proteins and expands them towards developing computational methods for characterisation of knot proteins and classification of knot proteins and unknot proteins. Published 278 known knot proteins have been used as the dataset in this investigation. The unknot dataset comprised of an equal set of hemoglobin proteins.

The present investigation has three distinct parts: Firstly, knot protein sequences were mapped into discrete signals using hydrophobicity values and their cross correlation was taken with discrete signals mapped from synthetic sequences of different hydrophobicity levels. This study revealed that a definite occurrence of hydrophobic domains exist in knot proteins. In the second part, the mapped sequences were spectrally analysed using Fast Fourier Transforms (FFT). Comparable spectral signatures were found for knot proteins and nonhub proteins (proteins which interact with less number of proteins) in the analysis. Its quantification led to the conclusion that knot proteins exhibit nonhub nature. The third part of investigation involved mapping knot protein sequences into a graphical representation known as Chaos Game Representation (CGR). Knot proteins exhibit special pattern in their CGR. These patterns were quantified by setting a ratio of CGR points clustering in the two opposite corners of their CGR (representing hydrophobic and hydrophilic amino acids).

The quantified parameters obtained from the above mentioned steps have been used as feature vectors to classify knot proteins and unknot proteins by ANN & Support Vector Machines. These feature vectors are used to locate knots in proteins.