11.1 CONCLUSION

This chapter gives the details of research carried out along with the limitations and recommendation for future enhancements. The first objective is to find an efficient loss less data compression algorithm, this goal is achieved by experimenting and comparing the preferences of various lossless algorithms like Huffman Encoding algorithm, RLE, Arithmetic Coding, LZ77 and Lempel-Ziv-Welch (LZW). From the experimentation and analysis LZW is found best, but the Computational Complexity of LZW is comparatively high. The LZW data compression algorithm is one of the most popular noiseless data compression algorithms for text and image sequences. The thesis presents novel approaches to reduce the computational cost and improve the performance of LZW. The conventional LZW algorithm requires huge computation when encoding, but the decoding algorithm is almost optimal when the dictionary is implemented with Linear Array. This enhancement is done using several phases. The first phase of this thesis reexamined some of the basic data structure implementations of LZW encoding like linear array with Linear search, Binary Search Tree (BST), Chained Hash table and also proposed a new basic LZW dictionary implementation using Binary Insertion Sort (BIS). The BIS is a novel sorting algorithm which is developed by exploiting the binary search.

The second phase of this thesis analyzed MDLZW architecture and Linear array, BST, Chained Hash Table and BIS is implemented and tested. The Multiple Dictionary (MD) architecture is very well suited for any basic data structure implementations or BIS implementation, and also this approach recues the limitation of BIS such as the average shift require after each insertion. From the experimentation and analysis MDLZW shows low computational cost when comparing with conventional LZW implementations.

In the third phase of this thesis is proposed a new data mining algorithm called Indexed K Nearest Twin Neighbour (IKNTN) Algorithm, the algorithm clusters and Indexes all possible twin elements. The algorithm is specially designed to optimize the performance of data structure. After clustering with the IKNTN the performance of data structures like BST, Linear array with linear search and Chained hash table is improved, and the computational complexity is reduced. So next stage in this phase is enhancing the LZW with
IKNTN, after enhancing with proposed IKNTN the computational cost is reduced when comparing with the previous implementations like Conventional LZW and MDLZW using any data structure and BIS implementation required few shift after each insertion.

The final phase of this research is done by clubbing MDLZW with IKNTN. The proposed optimization is done in two ways. First each and every single dictionary in MD is clustered using IKNT. Second each cluster is sub classified into multiple dictionaries. The proposed architecture gives optimal reduction of computational cost with any data structures like BST, Chained Hash Table, Linear array and BIS. The proposed algorithm may be the best architecture to reduce the computational complexity of LZW. The limitation of BIS LZW implementation is optimally reduced using this approach. The proposed chained hash table implementation algorithm achieved 17.47% low computational cost than MDLZW BST implementation, and the IKNTN_MDLZW chained hash table implementation found best when comparing with the experimented LZW methods in this thesis.

11.2 LIMITATIONS

✓ The speed of compression is taken as one parameter for comparing the performance.
✓ Only the twin neighbors are available in the same cluster.
✓ Euclidian distance is used for clustering.
✓ Multiple Index table require when clustering the dictionaries in MD.
✓ Possibilities of occurring Null dictionaries or empty dictionaries in the MDLZW.
✓ The time taken by BST implementation is increased for IKNTN_LZW, MDLZW_IKNTN and IKNTN_MDLZW.

11.3 FUTURE ENHANCEMENTS

The fixed M dictionaries set is used in the MDLZW that can be replaced with the variable dictionary set in order to avoid the Null dictionaries and space efficiency. Euclidian distance is used to find the nearest twin neighbour and clustering, the Euclidian distance can be replaced with some other distance measures to give more accuracy. Multiple index tables can be replaced with single index table to reduce the space complexity. Computational complexity of encoding is taken as one parameter. The proposed work can be further enhanced and expanded for the authentication of compression and decompression techniques to obtain optimum improvement based on the other parameters like compression ratio and space complexity.