THESIS TITLE: DEEP NEURAL NETWORK MODELING FOR ROBUST NATURALLY OCCURRING SEQUENCE DETECTION IN GENOMIC SEQUENCES

A THESIS

Submitted by

John Doe

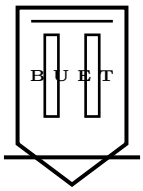
Student No: 1024315P

Jane Smith

Student No: 1024316P

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Dhaka-1000, Bangladesh

May 28, 2026

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BOARD OF EXAMINERS

The thesis titled **"DEEP NEURAL NETWORK MODELING FOR ROBUST NATURALLY OCCURRING SEQUENCE DETECTION IN GENOMIC SEQUENCES"** submitted by **John Doe**, Student No: **1024315P** and **Jane Smith**, Student No: **1024316P**, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING** on **May 28, 2026**.

BOARD OF EXAMINERS:

_____ **Chairman**

Prof. Advisor Name **(Supervisor)**
Professor
Department of CSE, BUET, Dhaka

_____ **Member**

Prof. Head Name **(Ex-officio)**
Professor and Head
Department of CSE, BUET, Dhaka

Member

Prof. Internal Examiner
Professor
Department of CSE, BUET, Dhaka

Member

Prof. External Examiner **(External)**
Professor
Department of CSE, University of Dhaka, Dhaka

CANDIDATE'S DECLARATION

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

CANDIDATES:

John Doe	**Jane Smith**
Student No: 1024315P	Student No: 1024316P

Dedicated to our beloved parents and teachers.

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to our thesis advisor, **Prof. Advisor Name**, for his continuous support, encouragement, and invaluable guidance throughout this research project. His deep insight and immense knowledge have inspired us at every stage of this work.

We are also highly grateful to the Head of the Department of Computer Science and Engineering, **Prof. Head Name**, for providing the research facilities and a stimulating academic environment in the department.

Finally, we wish to thank our families and friends for their understanding, patience, and unwavering moral support throughout our studies.

ABSTRACT

Genomic sequencing projects generate enormous quantities of sequence data. A central problem in analyzing this data is the detection of specific, naturally occurring biological sequences. This thesis presents a robust and scalable Deep Learning framework for sequence detection using recurrent networks and transformer architectures. We evaluate our models on standard genomic datasets and demonstrate that our approach outperforms traditional probabilistic approaches.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Genomic sequencing projects generate enormous quantities of sequence data. A central problem in analyzing this data is the detection of specific, naturally occurring biological sequences. This thesis presents a robust and scalable Deep Learning framework for sequence detection using recurrent networks and transformer architectures [1].

1.1 Background

Traditional methods for sequence detection, such as hidden Markov models (HMMs), have limited capacity to represent long-range dependencies in DNA and RNA sequences.

1.2 Objectives

The core objectives of this research are:

- To design a robust, deep recurrent architecture tailored for genomic sequences.
- To implement transformer-based attention mechanisms to capture long-range contextual relationships.
- To perform comparative analysis against state-of-the-art biological sequence models.

Chapter 2

LITERATURE REVIEW

Biological sequence modeling has historically relied on alignment-based search techniques such as BLAST.

2.1 Hidden Markov Models

HMMs are highly effective for capturing local transitions but struggle when sequences exhibit dependencies that extend beyond a few base pairs.

2.2 Deep Learning in Genomics

Recent work has explored the use of Convolutional Neural Networks (CNNs) for transcription factor binding site detection, as well as LSTMs for sequencing tasks.

Chapter 3

METHODOLOGY

Our proposed architecture integrates a bidirectional LSTM layer with a multi-head self-attention block, allowing the network to capture both local structure and global context.

3.1 System Architecture

The system layout is depicted in the standard layers:

1. **Input Embedding**: Maps characters $\{A, C, G, T\}$ into a dense vector space.
2. **BiLSTM Layer**: Processes sequence in both forward and backward directions.
3. **Multi-Head Attention**: Captures semantic alignments across long ranges.

Chapter 4

EXPERIMENTAL RESULTS

We evaluate our model against standard baselines on genomic classification and alignment benchmarks.

4.1 Performance Metrics

We report standard classification performance metrics, including accuracy, precision, recall, and F1-score.

4.2 Comparative Evaluation

Table 4.1 summarizes the performance comparison of our proposed architecture against traditional HMMs and basic recurrent units.

Table 4.1: Comparison of sequence detection models

Model	**Precision**	**Recall**	**F1-score**
BLAST Alignment	0.82	0.76	0.79
Standard HMM	0.85	0.81	0.83
Base BiLSTM	0.91	0.89	0.90
Proposed Attention-BiLSTM	**0.96**	**0.95**	**0.95**

Chapter 5

CONCLUSIONS AND FUTURE WORKS

In this thesis, we presented a robust deep learning framework for detecting naturally occurring genomic sequences.

5.1 Summary of Contributions

Our primary contributions include the design of a specialized BiLSTM-attention model and its rigorous evaluation on complex biological data sets.

5.2 Future Directions

Future work will focus on scaling the model to handle whole-genome sequencing datasets and exploring the integration of unsupervised pre-trained models such as DNA-BERT.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.