DISEASE TRAJECTORY MINING USING NETWORK ANALYSIS

Hira Riaz Haider
01-249212-003

Supervisor
Dr.Fatima Khalique

A thesis submitted in fulfilment of the requirements for the award of degree of Masters of Science in Data Science

Department of Computer Science

BAHRIA UNIVERSITY ISLAMABAD

2023

Approval of Examination

Scholar Name: Hira Riaz Haider
Registration Number: 75938
Enrollment: 01-249212-003
Program of Study: MS (DATA SCIENCE)
Thesis Title: Disease Trajectory Mining Using Network Analysis

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index 15%. that is within the permissible limit set by the HEC for the MS/M.Phil degree thesis. I have also found the thesis in a format recognized by the BU for the MS/M.Phil thesis.

Principal Supervisor Name: Dr.Fatima Khalique
Principal Supervisor Signature:
Date: 27-October-2023

Author's Declaration

I, Hira Riaz hereby state that my MS/M.Phil thesis titled is my own work and has not been submitted previously by me for taking any degree from Bahria university or anywhere else in the country/world. At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw/cancel my MS/M.Phil degree.

Name of Scholar: Hira Riaz

Date: 27-October-2023

Plagiarism Undertaking

I, solemnly declare that research work presented in the thesis titled "Disease trajectory mining using network analysis" is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me. I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS/M.Phil degree, the university reserves the right to withdraw / revoke my MS/M.Phil degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis.

Name of Scholar:Hira Riaz

Date:27-October-2023

Dedication

I dedicate this thesis first of all to MYSELF and to my father. Without his financial and emotional support the journey must have been different.

Acknowledgements

I was blessed to have supportive and friendly people who helped me preparing this thesis. I am thankful to my supervisor Dr. Fatima Khalique and co-supervisor Miss Mehroz who helped me wherever I was stuck during the research process. I am grateful for their critics, encouragement and guidance. My class fellows were tremendous to have a discussion on questions I had in my mind related to the research problem. I am thankful to my class fellows Misbah, Abu Bakr, Usama, Talha, Salman and Haider for their support.I am grateful to my fellow postgraduate student Anam, your continues support also helped me during the research process.

Abstract

Pakistan is struggling to provide good and on time healthcare facilities to its people.Situation become worse when the number of patients exceeds the capacity of available hospitals and healthcare facilities. With the help of current technology and data availability, we can analyse and forecast particular disease occurrence in the form of disease locations.This information is useful for healthcare sector and policy makers to make better decisions.This research analyse and forecast SARS and cholera disease locations for different cities of Punjab province.To conduct the study deep learning models LSTM(long short term memory) and Transformer models are used for forecasting dates and locations of SARS and cholera disease using 2016's year data set.

We believe that techniques used in this study to forecasting SARS and cholera disease dates and locations can be extended to other diseases as well.This will play an important role in improving the public health situation of Pakistan.Due to the data set availability we target different cities of Punjab province, however it can be extend to other province cities as well.

TABLE OF CONTENTS

AUTHOR'S DECLARATION ii

PLAGIARISM UNDERTAKING iii

DEDICATION iv

ACKNOWLEDGEMENTS v

ABSTRACT vi

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF SYMBOLS 1

1 INTRODUCTION 2
- 1.1 Background . . . 2
- 1.2 Problem Analysis/Formulation . . . 3
- 1.3 Research Objectives . . . 4
- 1.4 Significance of the study . . . 4
- 1.5 Scope Of Research . . . 4
- 1.6 Spatio Temporal Analysis . . . 5
- 1.7 Disease Trajectory . . . 5

2 RELATED WORK 7

3 METHODOLOGY 16
- 3.1 Spatio-Temporal Data . . . 16
- 3.2 Data Pre-processing . . . 16
- 3.3 Process Flow . . . 17
- 3.4 LSTM(Long Short Term Memory) . . . 18
  - 3.4.1 Cell State . . . 19
  - 3.4.2 Gates . . . 19

3.4.3 Output Gate 21
3.5 Transformer 21
3.5.1 Self Attention 22

4 ANALYSIS & RESULTS **24**
4.0.1 LSTM Results 25
4.0.2 Transformer Results 28
4.0.3 Model Performance 31

5 CONCLUSION & FUTURE WORK **33**

REFERENCES **34**

APPENDICES A-Y **38**

LIST OF TABLE

2.1 Literature Comparison 13

4.1 Data set features 24
4.2 Error comparison of LSTM and Transformer model 31
4.3 Model Performance Metrics Comparison 31

LIST OF FIGURE

1.1 Spatio Temporal Analysis 5

3.1 Process flow 17
3.2 Cells connected in a sequence in LSTM network 18
3.3 Internal structure of LSTM cell 19
3.4 Cell state of LSTM cell 19
3.5 Forget gate 20
3.6 Input gate of LSTM 21
3.7 Transformer workflow 22

4.1 Disease distribution 24
4.2 Monthly occurrence of SARS and Cholera 25
4.3 Weekly occurrence of SARS and Cholera 25
4.4 Line Plot of actual and predicted dates 26
4.5 Line plot of actual and predicted location 26
4.6 Confusion Matrix of Disease predicted by LSTM model 27
4.7 SARS trajectory of one week from 25 December till 31 December,where each color present individual day 27
4.8 Cholera trajectory of one week from 25 December till 31 December,where each color present individual day 28
4.9 Combined SARS and Cholera trajectory of one week.Circle is presenting SARS trajectory while triangle is presenting cholera trajectory 28
4.10 Line plot of actual and predicted Dates 29
4.11 Line plot of actual and predicted latitude and longitude 29
4.12 Confusion Matrix of Disease predicted by transformer model . . 29
4.13 SARS trajectory of one week from 25 December till 31 December,where each color present individual day 30
4.14 Cholera trajectory of one week from 25 December till 31 December 30
4.15 Combined SARS and Cholera trajectory of one week.Circle is presenting SARS trajectory while triangle is presenting cholera trajectory 30

LIST OF SYMBOLS

SARS – Severe acute respiratory syndrome

LSTM – Long short term memory

SGC – Spatial graph convolution

GCN – Graph convolution network

MLR – Multiple linear regression

KNN – K nearest neighbour

DTR – Decision tree regressio

GBR – Gradient boosting regression

XGBR – eXtreme Gradient Boosting regression

STNN – Spatio temporal neural network

RNN – Recurrent neural network

CHAPTER 1

INTRODUCTION

1.1 Background

The healthcare industry is critical to a nation's overall health. It is critical to have access to high-quality healthcare services such as medical treatment, preventative care, and health education. A well-functioning healthcare system guarantees that illnesses are recognized and treated as soon as possible, lowering the burden of chronic diseases and preventing epidemics.The entire quality of life is improved by having access to high-quality healthcare services. People may live healthier, more productive lives when healthcare is easily accessible and can meet the patient requirements. A productive population is one that is healthy. When individuals are healthy and have access to healthcare when necessary, they can actively participate in the labor force, which promotes economic growth.

Pakistan, since its independence, the government is struggling to provide better health care facilities to its people. In big metropolitan cities, the overall situation of health care sector is comparatively better than small cities.When it comes to health care facilities in rural areas, situation is worse.In the modern era of technologies, Artificial Intelligence has emerged as a powerful tool in solving many problems [1]. Health care sector can also be revolutionized with the help of AI [2].Machine learning is sub domain of AI, in which with the help of data we can solve particular problems.Prediction and forecasting comes under machine learning, in which we predict and forecast with the help of historical date.Deep learning is subset of machine learning which works on the basis of neural networks[3].Using deep learning techniques or models we can also forecast disease occurrence.

In a world where countries are striving to explore and discover new technologies, Pakistan is still facing issues to provide clean water and food to its people. Lack of clean water and food is infecting people with cholera.This thesis focuses on cholera and SARS disease forecasting.SARS, or Severe Acute

Respiratory Syndrome, is caused by a virus that spreads when someone who is sick coughs, sneezes, or talks, and tiny droplets with the virus get into the air. A lot of cholera and SARS patients are unable to have proper health services. Forecasting cholera and SARS disease locations can give a better understanding regarding spread of disease, to the health care sector to meet the requirements of patients in future.

1.2 Problem Analysis/Formulation

For successful public health planning and response, disease outbreaks must be understood and predicted. Disease trajectories, which combine geographical characteristics with a time horizon, can offer important insights into the dynamics of disease outbreaks. A lot of patients of Cholera and SARS disease are remained untreated due to lack of healthcare facility especially in disease outbreak. If we can predict the expected locations of Cholera and SARS day wise, this can help the government and health care sector to plan and provide proper health care facilities for the patients. This is the problem that this research set out to address.

Problem can be formulated as if we have historical data consisting of information regarding on which day disease occurred(time span) for a specific region using spatial features, we can forecast future disease locations and days.In this research, the disease-trajectory prediction can be formulated as predicting future trajectories of the target disease by using historical trajectories of the target disease. To facilitate the modeling of this problem, we firstly introduce some notations.The past trajectories of the i-th disease from day d = 1 to d =d_{obs} can be described by

$$X_i = \{t_i^1, t_i^2, ..., t_i^{d_{obs}}\} \tag{1.1}$$

where,

$$t_i^d = (x_i^d, y_i^d) \tag{1.2}$$

is the vector of disease i at day d, which includes the disease location coordinates, x as latitude and y as longitude. The future trajectories of the i-th disease can be expressed as

$$Y_i = \{t_i^{d_{obs}+1}, t_i^{d_{obs}+2}, ..., t_i^{d_{obs}+d_{pred}}\} \tag{1.3}$$

where,

$$t_i^{\tilde{d}} = (x_i^{\tilde{d}}, y_i^{\tilde{d}}) \tag{1.4}$$

are the predicted locations of disease i at day $\tilde{d}$

1.3 Research Objectives

The primary objectives of this research are as follows:

- To design a predictive model that combines both time series sequence and spatial features for forecasting disease dates and locations.
- To access model's reliability and accuracy using historical outbreak disease data
- To provide insights into the spatial and temporal patterns of disease outbreaks

1.4 Significance of the study

The research is based on Time Series Analysis and Forecasting of cholera and SARS disease. Disease forecasting gives early warnings about possible epidemics. This helps public health professionals and governments to prepare and deploy resources, such as medical supplies, healthcare workers, and facilities, to respond effectively to the epidemic.By predicting where and when both diseases might spread, government can strategically distribute medical supplies, medications, and healthcare workers to areas at higher risk.Forecasting helps plan healthcare capacity. Hospitals and healthcare facilities can prepare for potential surges in patient numbers and allocate beds, ventilators, and other medical equipment accordingly.Forecasting informs policy decisions. Government can make evidence-based decisions about reopening, lockdowns, travel advisories, and other policy measures based on forecasted disease trends.

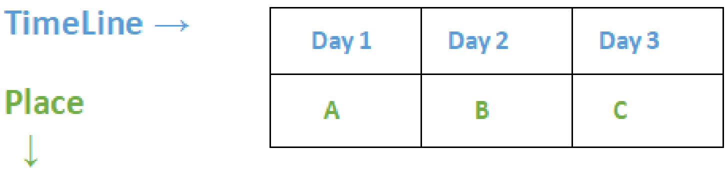
1.5 Scope Of Research

The target group of the research are researchers interested in finding disease spread using deep learning techniques through spatio temporal analysis using historical time series data.In addition, it is also relevant for public health officials and epidemiologists. This research can facilitate decision-making processes related to disease surveillance, resource allocation and the implementation of preventive measures for Cholera and SARS.

1.6 Spatio Temporal Analysis

location and time span two basics aspects of spatio temporal analysis[4].Spatio-temporal analysis involves looking at two crucial factors at the same time: ”where” and ”when.” It’s similar to viewing two aspects at the same time, namely the spatial aspect, which defines locations such as cities, and the temporal aspect, which denotes moments in time over various days. Consider the following scenario: we have a map that represents various places and a timeline that defines distinct time points over different days. The use of this integrated map and timeline to acquire insights into the dynamics of events, understanding how they change and traverse both space and time, is known as spatio-temporal analysis.Timelines can include accurate timestamps down to the minute or second, or they can be aggregated to daily, weekly, or monthly intervals in some circumstances. When we aggregate data to daily intervals and examine it on a daily basis, we are still engaging in spatio-temporal analysis.

Here’s a simple diagram to understand it better.



Day 1	Day 2	Day 3
A	B	C

Figure 1.1: Spatio Temporal Analysis

In this diagram, the map represents different places (A, B, C), and the timeline represents different days (Day 1, Day 2, Day 3). With spatio-temporal analysis, we can answer questions like:

”Where did events happen on Day 2?” ”How did the situation change from Day 1 to Day 3 in Place C?” By combining both space (where) and time (when) information, we can see patterns, trends, and relationships that we might not notice if you only looked at one thing at a time.

1.7 Disease Trajectory

A set of geographic data points that indicate the location of the item at various times in time are frequently used to illustrate trajectories. Latitude, longitude, altitude, and occasionally other characteristics like speed or direction, can all be included in these data points as coordinates.Insights into how objects or entities move and interact with their surroundings through time are provided by trajectory data, which is useful for spatial analysis.We now have a significant number of GPS trajectories capturing people's position histories because to the growing number of GPS-enabled gadgets that are available[5]. Using trajectory analysis, one may get a deeper comprehension of spatial dynamics, patterns, and trends that can then be used in a variety of contexts, such as transportation, ecology, urban planning, and more.

Disease trajectory is like tracing the path of a disease as it moves and spreads. It's similar to following the steps of a journey. Imagine we're drawing a map that shows where a disease, like cholera and SARS, starts and how it moves to different places over time.

For example, let's think about cholera spreading in different cities:

Starting Point: The disease might begin in one city. Movement: As time goes on, the disease might move to nearby cities or other areas. Spread: More and more cities might get affected as the disease spreads. Locations: We can predict where the disease might show up next based on how it has been moving. By studying the disease trajectory, we can figure out where the disease might go next and help plan ways to prevent it from spreading further. It's like drawing a map that helps us see where the disease is headed and how to stop it.

CHAPTER 2

RELATED WORK

Many researchers have conducted studies on the movement or trajectories of individuals and objects to have better intelligent transportation systems. Pedestrians are one of the main components of intelligent transportation systems. Understanding how they move helps in a variety of fields such as self-driving automobiles, robot navigation, and smart tracking systems. Self-driving cars are improving, and understanding where people will walk is critical for them. It assists the automobiles in determining whether to stop, slow down, turn, or continue driving. Unlike automobiles, pedestrians may move in a variety of directions, which can be challenging for self-driving cars, particularly in areas with a high density of people and cars, such as neighborhoods, marketplaces, or parking lots. Predicting where people will walk can also aid with traffic management.A convolutional long short-term memory (LSTM) neural network architecture is used for pedestrian trajectory prediction in intelligent transportation systems [6].

To forecast where pedestrian would walk or future pedestrian trajectory, the authors developed a system called spatial graph convolutional LSTM (SGC-LSTM). First by using LSTM,it determine how each individual has been moving. Then, they utilize a specialized network known as graph convolution network(GCN) to observe how individuals interact with one another in a group. They integrate these movement and interaction elements, and an emotion gate determines how important the interactions are. Finally, another LSTM is used to predict where pedestrian would go next.

The majority of earlier models either used max-pooling techniques to extract certain characteristics from movement hidden states in order to capture pedestrian interactions, or they simply fused movement hidden states. These policies produce incomplete information on interactions because they cannot find correlations inside the concealed states.For machine learning on graphs, a potent neural network design is the graph convolution network [7][8]. By using message passing between the nodes of graphs with an adjacent matrix, GCN can represent the importance of graphs [9].Not only in recommend systems [10] but also in traffic speed forecasting GCN(Graph Convolutional Network) has played a pivot role[11].

For applications requiring sequence prediction, RNNs and its variant structures—such as LSTM[12] networks and gated recurrent unit (GRU) [13]are used in many domains.

For trajectory prediction attention base models have been also used. The encoder-decoder model based on RNN can perform better due to attention. Its major goal is to assist models in extracting more vital and significant information by giving each component of the input a distinct weight. When compared to a standard RNN, the attention mechanism helps the model to focus on the aspects that are important for prediction while disregarding the aspects that are not significant, enabling the model to reach more accurate conclusions. Numerous attention-based techniques have been employed to assess the relative significance of pedestrians in trajectory prediction [14][15][16][17]

A Spatial-Temporal Attention and Graph Convolutional Network Model (STAGCN) framework is helpful in anticipating urban transportation destinations, which is an important part of urban traffic planning and congestion management[18]. To extract latent, low-dimensional representations from complicated multi-graph input data, this novel model leverages a Graph Convolutional Network (GCN) with two hidden layers. GCN produces time series data, which is then fed into an LSTM (Long Short-Term Memory) encoder model to capture temporal relationships within the data.

To learn spatial and temporal weight vectors,the hidden outputs of the LSTM encoder at various time points are then fed into an attention module designed. Then in LSTM decoder,the weighted spatial-temporal vectors are sent which is in charge of forecasting future destination.For modeling temporal connections within location sequences LSTM encoding module is used, whereas the decoding module is responsible for predicting future locations. It includes aspects like geographical connection between sites and their surroundings, . This method recognizes that relying exclusively on LSTM for predictions may be insufficient. For improving the accuracy of destination predictions,.the addition of a spatial-temporal attention mechanism enables efficient collection of temporal and spatial data, as well as their interaction. Author has considered short intervals for forecasting, for longer intervals, when the network scale increases, getting better results needs to be explore.

Just like predicting traffic and pedestrian trajectory with the help of time series data, we can also predict and forecast disease or patient trajectory.The movement of the patient defines the disease trajectory.When pandemic COVID-19 hit us, a lot of researcher focused on study and predict its spread. [19] addresses movements of pandemic disease patients, COVID-19 patients in Nigeria.Six different machine learning models MLR(multpile linear regres- sion),KNN(K nearest neighbor), DTR(Decision tree regression),RFR(Random forest regression), GBR(Gradient Boosting Regression), and XGBR(eXtreme Gradient Boosting regression) are used for predicting patients locations. A sample of 15 people drawn from the Local Government Area of Lagos State, Nigeria for collecting location records.These Trajectory dataset contain geographical information acquired in real-time from mobile devices tablets and smartphones and, such as altitude, latitude, longitude, time, and more.By placing a tracking App on each participant's individual smartphone to enable experimental contact tracing, and all participants readily consented to participate in the study.In comparison to MLR, kNN, DTR, GBR, and XGBR, the DTR algorithm predicts the next probable location of a COVID-19 patient more accurately. Due to privacy issue less patients participated to collect real time patients locations. More relevant features like town or nearby area etc should have also used to predict more accurate locations.

Another study proposes spatio-temporal neural networks for forecasting spread of Covid-19 [20].The Spatio-Temporal Neural Network (STNN) successfully incorporates temporal and spatial characteristics, outperforming traditional epidemic forecasting models in terms of accuracy and how it handles spatio-temporal data.STNN with Augmented Spatial States (STNN-A) and STNN with Input Gate (STNN-I) are two augmented Spatio-Temporal Neural Network (STNN) architectures introduced by the authors. These architectural changes increase the predictability and flexibility of the STNN models used to estimate COVID-19 results.

The STNN model are compared to standard epidemic prediction models such as the fully-connected neural network , recurrent neural network (RNN), traditional curve fitting models. The results of numerical simulations show that the STNN models outperform in both fitting and prediction tasks.The study collects Chinese province data from the People's Republic of China's National Health Commission, US data from Johns Hopkins University, and Italian data from the Health Ministry. These files provide the total number of verified COVID-19 cases, deaths, and recoveries.

In addition to patient numbers, the study collects temporal and regional information about the disease's propagation. This includes daily migration statistics across Chinese provinces, hospital-related information (such as the number of hospitals and fever clinics), and meteorological data that includes temperature, humidity, and air quality index measures.One of the main limitation is that using very complicated polynomial curves for fitting data can be risky. They might not work well outside the range of the data we have. So, one has to be very careful when trying to make predictions or estimates using these kinds of curves.

Germany also conducted many researches regarding spread of COVID-19 like China.One of the study examine the COVID-19 weekly spread[21].The research study proposes a multimodal learning approach for predicting local COVID-19 instances in Germany that includes graph neural networks (GNNs). Research presents a new fusion approach that directly merges mobility and connection data from the online platform Facebook with structural and geographical features of cities and districts across Germany. The network learns the embedding of each district autonomously from start to finish, removing the requirement for a separate preprocessing phase.This aligns with prior suggestions advocating for the adoption of flexible and hybrid methods that consider collaborative sources of information.

From a methodological standpoint, they have achieved integration by merging graph neural networks with epidemiological models. This enables to handle both network-based data and tabular data concurrently. Moreover comparisons, validation assessments, and quantification of uncertainty to evaluate the trustworthiness of our model is considered. The model gives predictions for weekly COVID-19 cases at the community level, covering all 401 federal districts in Germany.This strategy presents an extensive and adaptable approach for forecasting COVID-19 cases, capitalizing on a diverse array of data sources and advanced modeling methods.They divide data sources into infection data and Facebook data relating to human movement and connectivity. As time series data, the infection data only contribute to the model's structured and target components. In contrast, the majority of network data is used directly within the GNN module. They transform the networks into units consistent with our time series data to enable validation tests and coefficient interpretation. This is accomplished by extracting various structural properties from the networks.

To study long term spatio temporal dependencies for the spread of influenza like illness(ILI) GNN model is used[22].The major focus of this research is on the long-term prediction of Influenza-Like Illness (ILI) patient numbers spanning 2 to 15 weeks. To address the complexities of research, combines graph propagation models with deep spatial representations, with the goal of compensating for the dataset's intrinsic lack of temporal information. Considering each geographical place to be a node, resulting in a well-crafted graph neural network architecture.This framework has been carefully designed to serve as a guide for understanding how illnesses spread throughout people.Researcher also utilized enlarged temporal convolutions, a deliberate choice that helps to spot patterns in the data that occur across both short and long time periods.

Another research [23] demonstrates the usefulness of comprehensive mobility data,for tracking the disease spread using spatio-temporal graph neural network developed for COVID-19 prediction at the county level in the United States. Authors recommend using more detailed real-time data that can capture regional movement and inter-region interactions to improve forecast accuracy.Considering future illness cases in a location are impacted not just by past data from that region, but also by data from other places, such as persons going to and from that region and regions with comparable epidemic trends.

The increasing use of GPS-enabled mobile devices provides a fresh and extremely reliable source of mobility data, otherwise to capture regional movement data has been difficult.In addition, they suggest a unified method to infectious illness modeling that can account for both temporal and geographical interactions.The authors argue that by including inter-regional connections and mobility data, their suggested technique employing Graph Neural Networks may effectively capture complex dynamics and improve the accuracy of COVID-19 case predictions. For a given time series serving as input and the model predicting either a single value or a future time series, this can be framed as a regression learning job.

However, this technique is not suitable for analyzing human motion movement, data is naturally represented as a spatial network, edge weights representing estimates of human movement between locations and with each node representing a place or location linked to many other nodes.To capture both geographical and temporal connections, a graph with many edge types can be used.Edges in the spatial domain represent direct movement between sites and are weighted depending on normalized mobility flow. To represent the county connectivity graph on a particular day, the network structure is made of 100 stacked layers.To capture full week temporal information,inside each layer each node makes direct links with the seven nodes that come before it in time.By using this approach, both geographical and temporal aspects are captured, which allow the model to give exact forecast for COVID-19 patients.The New York Times (NYT) COVID-19 dataset,gives information about the number of COVID-19 cases in different regionsThe dataset which helps the researchers understand the quantity of movement in different regions, Google COVID-19 Aggregated Mobility Research Dataset was used. This dataset provides information on the number of trips taken by people in different regions, which can be useful in predicting the spread of the virus.There was another dataset which was used, Google Community Mobility Reports, which provides information on the dynamics of various types of movement. This dataset provides information on the types of places people are visiting, such as workplaces,parks, grocery stores.This information can be useful in predicting the spread of the virus and understanding how people are responding to the pandemic.However, The information is only available to smartphone users who have enabled Google's Location History function, which is disabled by default. As a result, the statistics may not be representative of the full population and may differ depending on location. The data cannot be utilized to establish direct comparisons between areas because areas differs significantly.The data does not capture significant Other factors that can affect COVID-19 transmission, like changes in the weather and increased prevalence of wearing masks. These factors, combined with increased awareness, can effectively reduce transmission even when mobility remains unchanged.These external features should be considered for accurate forecasting.

Many studies have shown that using transformers gives best forecasting. One of the study [24] uses transformers to forecast influenza disease.For protecting and preventing outbreak, researcher focuses on developing transformer base model on the dataset to capture long range dependency.The transformer model is compared with the autoregression model and RNN model. To improve the accuracy of the predictions, the researchers have designed a sources selection module that measures curve similarity to fuse information from different data sources and capture spatial dependency.To improve the accuracy of the predictions, the researchers have designed a sources selection module that measures curve similarity to fuse information from different data sources and capture spatial dependency.To handle long range sequence of time series data,input selection module and the encoder decoder module is used.

Most researchers have studied the dynamics and spread of a single disease. While in real life scenario where there is outbreak of more than one disease, these needs to be addressed and studied. So there is need to study,analyze and predict the spread of multiple diseases at the same time.

Table 2.1: Literature Comparison

Ref	**Description**	**Model**	**Data Set**	**Shortcomings**
[6]	Social graph convolutional LSTM for pedestrian trajectory prediction	GCN and LSTM	Pedestrian trajectory dataset	In real world scenario the mobility of pedestrian does not only depend on only neighbours pedestrian within a specific distance, proposed methodology does not assumes this.
[18]	Spatial-Temporal Attention Mechanism and Graph Convolutional Networks for Destination Prediction	GCN and LSTM	Large scale urban real dataset	They have considered short intervals for forecasting, for longer intervals, when the network scale increases, getting better results needs to be explore.
[19]	A framework for monitoring movements of pandemic disease patients based on GPS trajectory datasets	MLR, KNN,DTR, RFR, GBR, XGBR	covid-19 patients trajectories dataset	Due to privacy issue less patients participated to collect real time patients locations. More relevant features like town or nearby area etc should have also used to predict more accurate locations.

[20]	Spatio-Temporal Neural Network for Fitting and Forecasting COVID-19	Spatio-Temporal Neural Network	Chinese Covid-19 patients dataset	Very complicated polynomial curves for fitting data can be risky. They might not work well outside the range of the data they have. So, one has to be very careful when trying to make predictions or estimates using these kinds of curves.
[21]	Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly COVID-19 cases in Germany	GNN	Covid 19 data provided by Robert–Koch Institute for Germany	By adopting time-varying networks, extending semi-structured approaches to include epidemiological models, and exploring additional data sources with finer granularity may enhance spatio-temporal models
[22]	Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction	GNN	Infectious Disease Weekly Report (IDW)	need to explore complex dependencies like social factors, climate changes, and population migration and simultaneously analyze multiple locations to identify key factors driving epidemic outbreaks

[23]	Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks	GNN	New York Times COVID-19 dataset, Google COVID-19 Aggregated Mobility Research Dataset	External factors like change in weather and use of masks are not considered, which will help in accurate foracasting.
[24]	Long-term prediction for temporal propagation of seasonal influenza using Transformer-based model	Transformer	Japan and US infleunza patients weekly dataset.	Weather conditions like wind speed, humidity and temperature should be added to the model to improve the accuracy

CHAPTER 3

METHODOLOGY

3.1 Spatio-Temporal Data

Spatial features refer to the representation of geographic locations, typically using latitude and longitude coordinates. Disease spatial features, in particular, indicate precise disease locations. When these disease locations are recorded in relation to a specific time period, they can form a sequence that shows how the disease spreads over time.

To create this sequence, we use dates to establish a timeline. By organizing disease locations day by day, we can observe the progression of the disease. In our study, we've utilized latitude and longitude to represent the spatial aspect (geographic locations), while the temporal aspect (time-related information) is captured through the dates of disease occurrence. This approach helps us analyze how the disease moves through different locations over time.CNN, RNN, GRU's are used in many spatio-temporal analysis, but as LSTM and transformers are better to capture long range dependency and sequence, we have used LSTM and transformer model to perform sptio-temporal analysis.

3.2 Data Pre-processing

As deep learning models typically work with numerical data rather than text representations, it's necessary to preprocess the data set accordingly. In our data set, dates are initially in text format. To make them suitable for modeling, we first convert them into a date time format and then arrange them in a day-wise manner to create a time series data set.

To represent these dates in a format that preserves their temporal sequence, we have used ordinal encoding. This encoding method assigns a unique sequential number to each date based on its position in the chronological order, creating a clear sequence of dates.

Additionally, to differentiate between different disease categories, we utilize integer encoding. In our study focusing on SARS and cholera diseases, integer 0 can represent SARS and the integer 1 to represent cholera. This encoding simplifies the categorization of diseases within the data set and facilitates subsequent modeling and analysis.

3.3 Process Flow

To predict the occurrence of SARS and Cholera diseases along with their respective dates and locations, our process begins with a crucial step: data pre-processing. In this stage, we convert the dates into a sequential format, and the disease feature is transformed into categorical data. After cleaning and preparing the data, we partition it into training and testing subsets.

Before input data to LSTM model, it has to be reshape into 3-dimensional data.First dimension present the batch size, second dimension sequence length in the input data and last dimension number of input variable.For transformer model, input data has te be reshape into fixed length sequences which is also known as window size. After reshaping the data, it is given as input to the respective model.To evaluate the performance of our models, both LSTM and Transformer, we conducted a thorough comparison of the results. This comparison involved visual inspections of the predictions as well as the utilization of evaluation metrics.Below figure 3.1 shows process flow for this research.

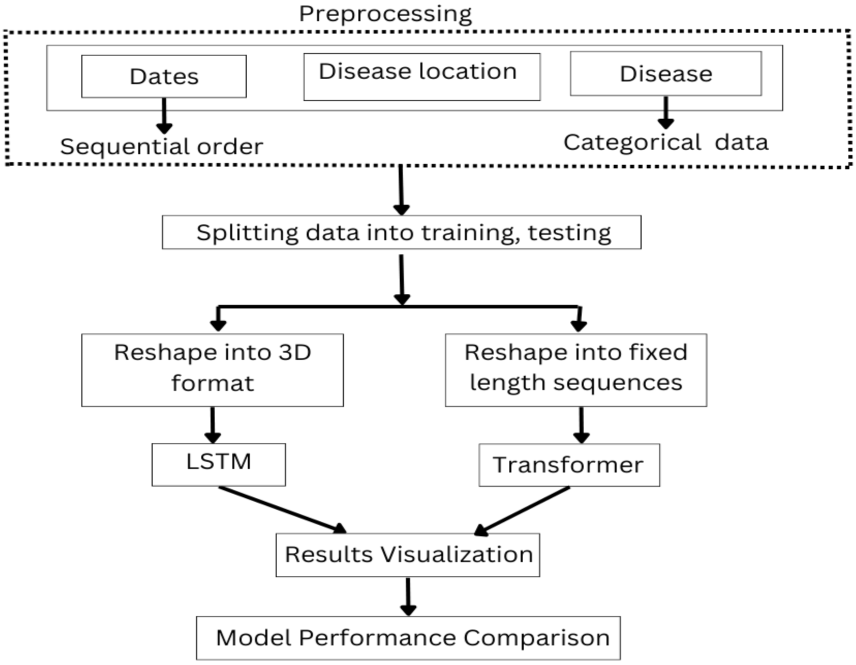


Figure 3.1: Process flow

3.4 LSTM(Long Short Term Memory)

LSTM(long short term memory)is a type of RNN capable of long term dependencies[25].LSTM is basically a series of cells or block which are connected in sequence. Below is the figure of a simple LSTM network which is consisted of three cells connected in a sequence.

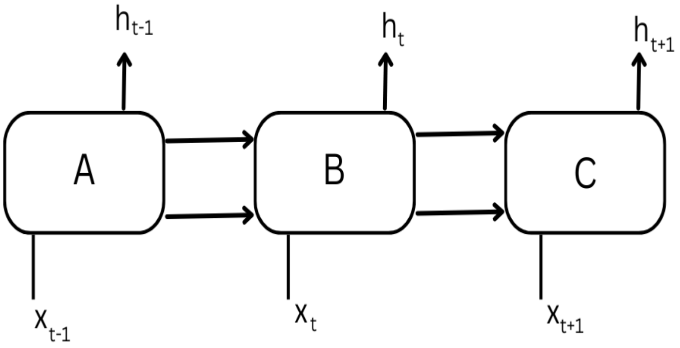


Figure 3.2: Cells connected in a sequence in LSTM network

$t-1$:Previous day
t:Current day
$t+1$:Next day
H_t-1:hidden state of the previous day
H_t: hidden state of the current day
H_t+1: hidden state of the next day
X_t-1:input of previous day
X_t:input of current day
X_t+1:input of next day

The first cell A passes two informations to the next cell which is presented by double arrow. The two informations which are passes to the next cells are cell state and the hidden state. Each cell has an input which is represented with X and output is the hidden state denoted denoted by h.

X_t-1 means input of previous cell state, H_t-1 means hidden state of the previous day.X_t is the current input,H_t current hidden state.X_t+1 input of next cell and H_t+1 is the next cell hidden state. The forward arrows shows the flow of information(cell state and hidden state) from one cell to another.

Let us consider a single cell in more detail,how a single cell process the data and what are the main parts of a single LSTM cell.The below figure shows the internal structure of LSTM cell.

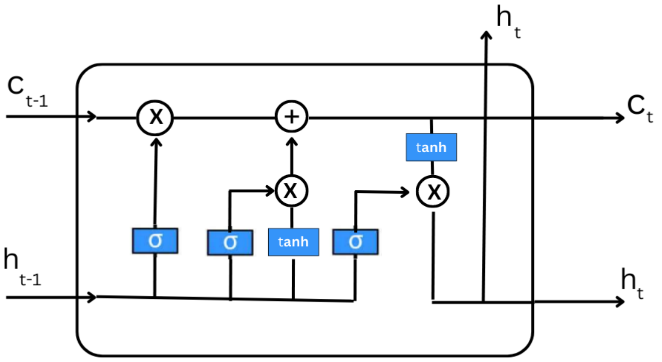


Figure 3.3: Internal structure of LSTM cell

The cell takes input of the current day denoted by X_t along with the cell state of the previous day denoted by $C_t - 1$ and hidden state of the previous day denoted by $h_t - 1$.Let us consider cell state first to better understand how it is connected with other parts.

3.4.1 Cell State

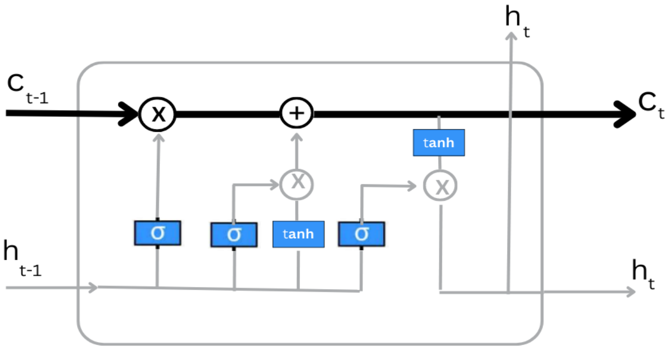


Figure 3.4: Cell state of LSTM cell

The cell state is represented by the horizontal line at the top as shown in above figure, and it palys a role of conveyer belt, moving data linearly through the data channel. It is relatively easy for information, to just travel across each LSTM cell unmodified.

3.4.2 Gates

The LSTM may delete or add information to the cell state, which is carefully controlled by structures known as gates. Gates are a method of optionally allowing information to pass through. They are made up of a layer of sigmoid and a point-wise multiplication process. The sigmoid layer produces integers ranging from zero to one, indicating how much of each component should be allowed through. A value of zero indicates "let nothing through," whereas a value of one indicates "allow everything through!". LSTM has three gates, forget gate, input gate and output gate.

These gates helps the LSTM to protect and have a control on the state of cell.All these gates along with the cell state is what make LSTM capable of process the information sequentially. Starting from forget gate we will go step by step to have a better understanding how data is processed in a cell.

Forget Gate

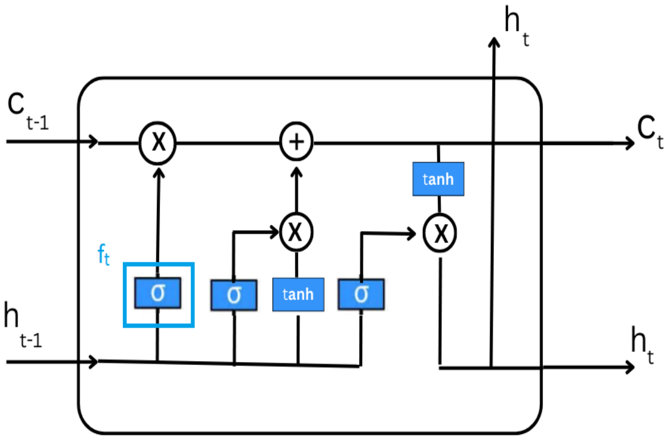


Figure 3.5: Forget gate

In the above diagram, blue highlighted box is what makes forget gate.This gate controls whether information should be discarded or saved. The sigmoid function is used to process information from the prior hidden state as well as information from the current input.The values range from 0 to 1. The closer to 0 , the more to forget, and the closer to 1, the more to keep. .Let forget gate be denoted by f_t below is the equation for forget gate.

$$f_t = \sigma(X_t * U_f + H_t - 1 * W_f) \tag{3.1}$$

X_t: input to current day
U_f: weight associated with input
$H_t - 1$: hidden state of the previous day
W_f: weight matrix associated with hidden state
σ:sigmoid activation function

Input Gate

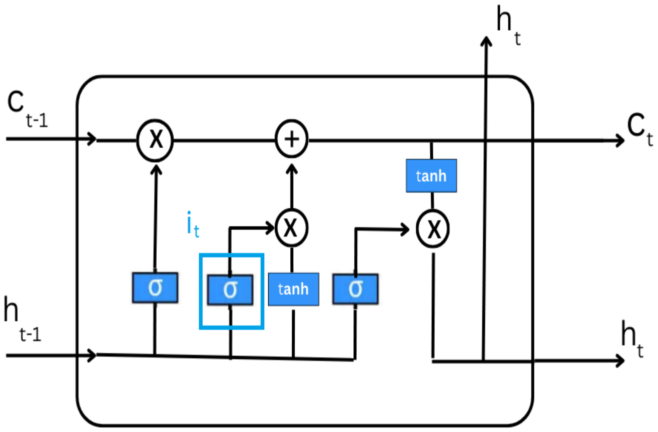


Figure 3.6: Input gate of LSTM

Input Gate updates the cell state. The input equation gate can be given as below.

$$i_t = \sigma(X_t * U_i + H_t - 1 * W_i) \tag{3.2}$$

X_t:input to current day
U_i:weight matrix of the input
$H_t - 1$: hidden state of the previous day
W_i: Weight matrix of input associated with hidden state
σ:sigmoid activation function.

3.4.3 Output Gate

Equation for output gate can be given as below

$$O_t = \sigma(X_t * U_o + H_t - 1 * W_o) \tag{3.3}$$

Output gate calculates the value for the next hidden state.Let output gate denoted by O_t

3.5 Transformer

In the field of NLP, such as machine translation, sentiment analysis, and text production, transformer networks have had considerable success[26].Transformer networks utilize the well-known encoder-decoder structure found in RNN seq2seq models [27][28]. The diagram shown below is an overview of transformer architecture encompassing two main modules encoder and decoder.

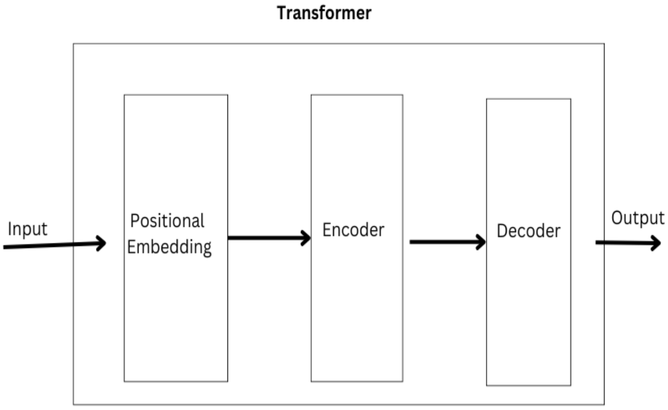


Figure 3.7: Transformer workflow

In the initial step embedding is performed which consists of tokenization and positional embedding , these embeddings are given as input to the encoder module.Positional embeddings are added to the token embeddings to offer information about the relative or absolute location of the tokens in the sequences because the transformer model lacks an intrinsic understanding of the order of tokens. To the token embeddings are vectors known as positional embeddings. Positional embeddings may be created in a variety of methods, but one popular method is to utilize sine and cosine functions with various frequencies.By combining the positional embedding with the token embedding, the final embedding for each token is achieved. The first layer of the transformer model is then fed these final embeddings as its input.Along with the other model parameters, the embedding matrices (for tokens and locations) are learnt during the training procedure.

3.5.1 Self Attention

Both the encoder and decoder module have self attention mechanism. Each layer in a typical neural network takes an input with a predetermined size and converts it into an output. However, in many real-world situations, the input data is sequential and variable in length (for example, a phrase or a document etc), and the output at each place could be influenced by the complete input sequence.This problem is solved by the self-attention mechanism, which enables the model to consider the importance of each component of the input sequence while processing a certain place in the sequence.In order to capture long-range relationships in the data, the self-attention mechanism enables the model to concentrate on various portions of the input sequence while processing each sequence.A "head" is a distinct self-attention mechanism in the transformer model. To process the input sequence in several subspaces, the transformer model runs many heads concurrently. The term "multi-head attention" describes this. The concept is to have numerous sets of Query, Key, and Value weight matrices as opposed to a single set, and to apply the self-attention mechanism differently to each set. The final output is created by concatenating and linearly transforming the outputs of each head.Transformer's central concept is to totally replace repetition with a multi-head self-attention mechanism.

The self attention mechanism of transformers first learns the query matrix Q, key matrix K and a corresponding value matrix V, for all embedding of sequence spanning from t to T.For dimension of each query d_k,attention equation can be given as:

$$Att(Q,K,V) = \frac{Softmax(QK^T)V}{\sqrt{d_k}} \tag{3.4}$$

The attention equation in the Transformer model, denoted as Att(Q, K, V), computes the weighted combination of values (V) based on the similarity between the query (Q) and the keys (K). This similarity is determined by the dot product of Q and K, divided by the square root of the dimension of the keys, and then softmax is applied to obtain the attention weights. Essentially, it's a mechanism that allows the model to focus more on relevant information when processing sequences, making it a fundamental component of the Transformer's success in sequence to sequence tasks.

CHAPTER 4

ANALYSIS & RESULTS

This chapter discusses the data set used for performing this research, along with results analysis and discussion.Data set of SARS and cholera disease for different locations of Punjab,Province is studied for this research.Data set has four features date, latitude, longitude and disease. Date feature specify the day of disease location. Latitude and longitude features present geographical location, while disease feature specify disease, it can be SARS or cholera.The below table shows features for this study.

Table 4.1: Data set features

Date
Latitude
Longitude
Disease

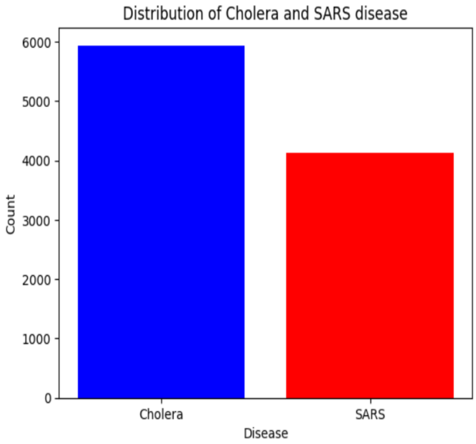


Figure 4.1: Disease distribution

If we look at the distribution of disease, 4127 SARS cases and cholera 5940 cases were recorded.Figure 4.1 shows bar chart for the disease distribution.

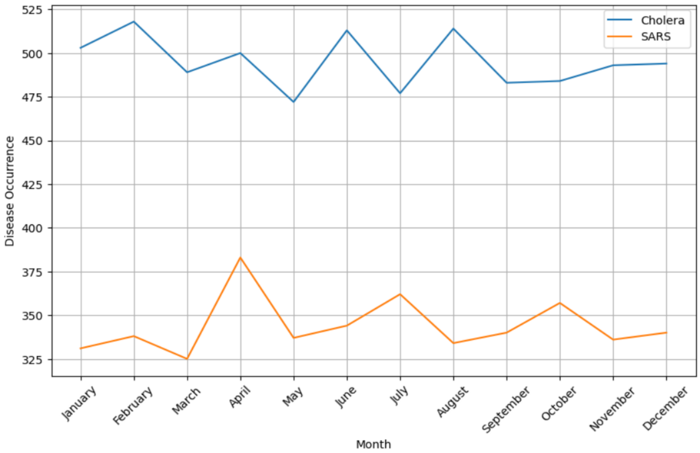


Figure 4.2: Monthly occurrence of SARS and Cholera

The above figure shows the monthly occurrence of SARS and cholera disease.There are more number of cholera cases in the month of February, however number is decreasing from September on wards. High number of SARS occurred during the month of April. The below figure shows weekly occurrence of both SARS and Cholera.

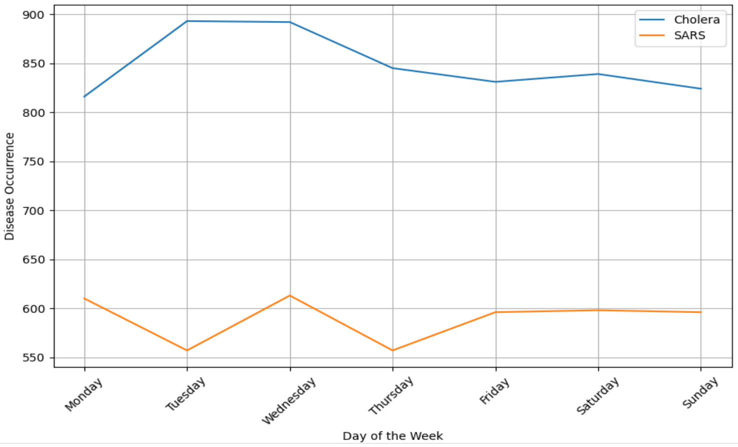


Figure 4.3: Weekly occurrence of SARS and Cholera

4.0.1 LSTM Results

Using LSTM model we are predicting disease dates and locations. The figure below shows next 10 sequence(index) actual and predicted dates. As dates are ordinal encoded to capture the sequence of day, below line plot shows encoded dates on y axis and Index as x-axis.

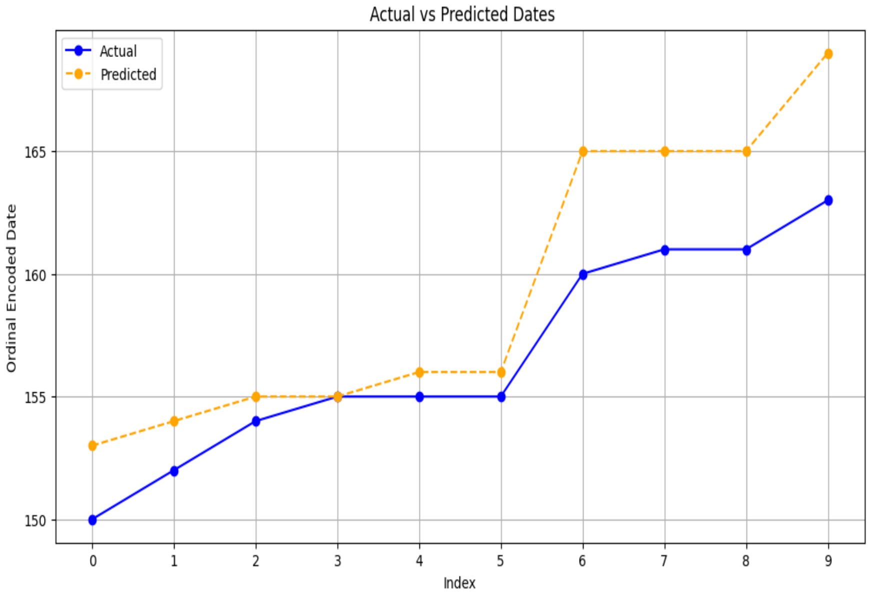


Figure 4.4: Line Plot of actual and predicted dates

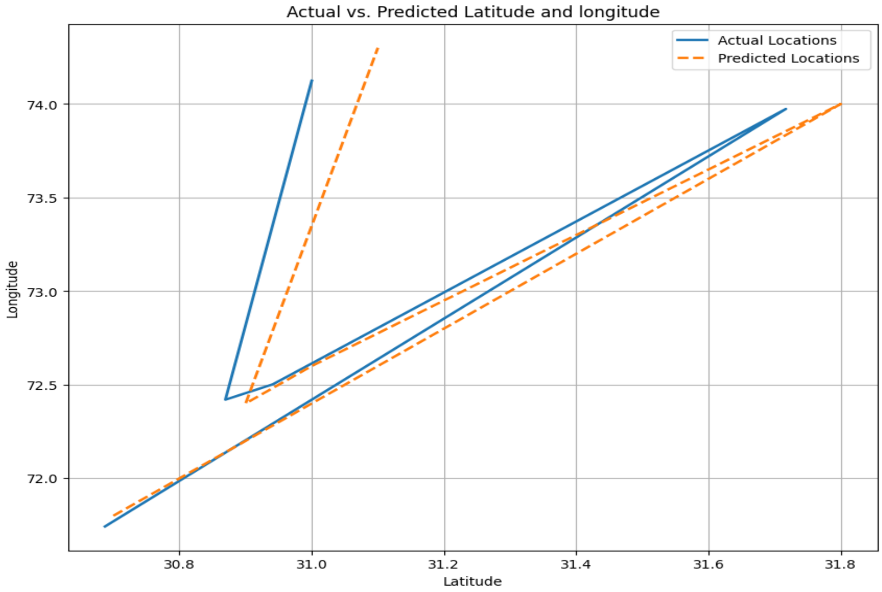


Figure 4.5: Line plot of actual and predicted location

Latitude and longitude can be combined to specify a location which can be presented as line plot.Figure 4.6 shows the actual and predicted latitude and longitude. Since the data is distributed in a manner that 80% is used for the training and rest of 20% for testing. The testing part consisted of almost 20000 data points, to visualize all those 20000 predicted latitude and longitude, it resulted in cluttered line plot of actual and predicted latitude and longitude.To avoid this, we have taken 30 actual and predicted latitude and longitude.

We can calculate correct predictions of disease using confusion matrix.With the help of confusion matrix, we can check how many SARS and Cholera disease are predicted correctly. The below figure shows confusion matrix for disease feature using LSTM model.

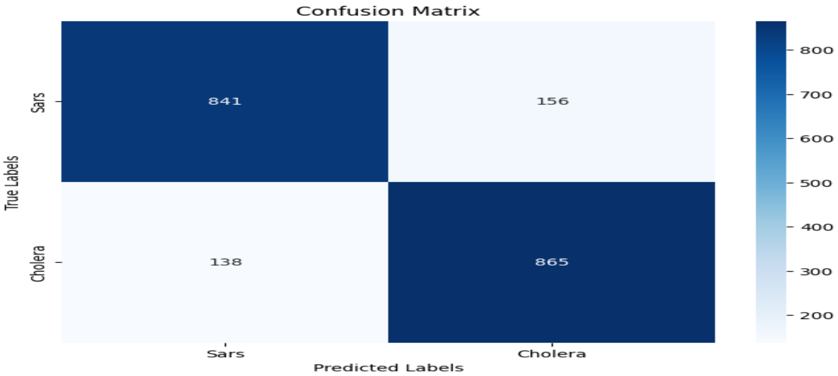


Figure 4.6: Confusion Matrix of Disease predicted by LSTM model

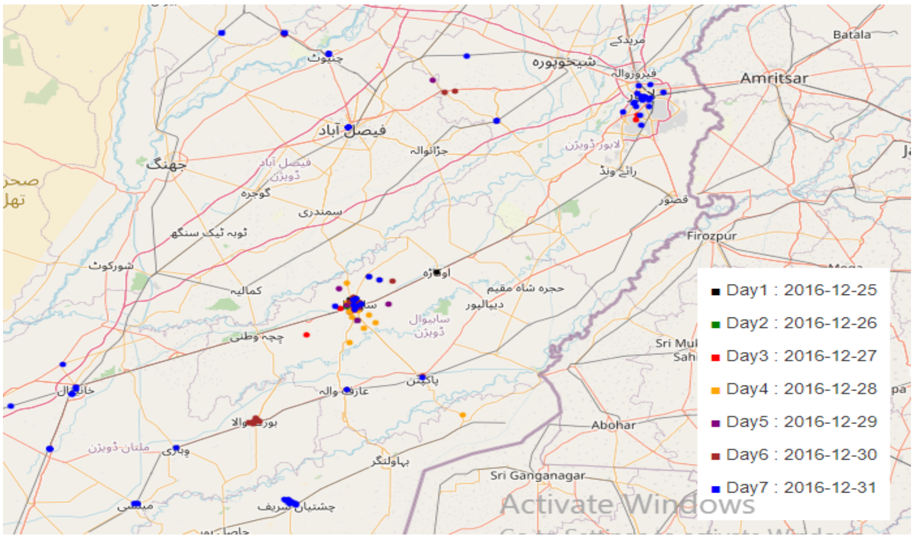


Figure 4.7: SARS trajectory of one week from 25 December till 31 December,where each color present individual day

Figure 4.7 and 4.8 shows one week trajectory of SARS and cholera on map, where figure 4.9 shows a combine trajectory of SARS and cholera.Circle is showing SARS trajectory while a triangular shape to show cholera trajectory.

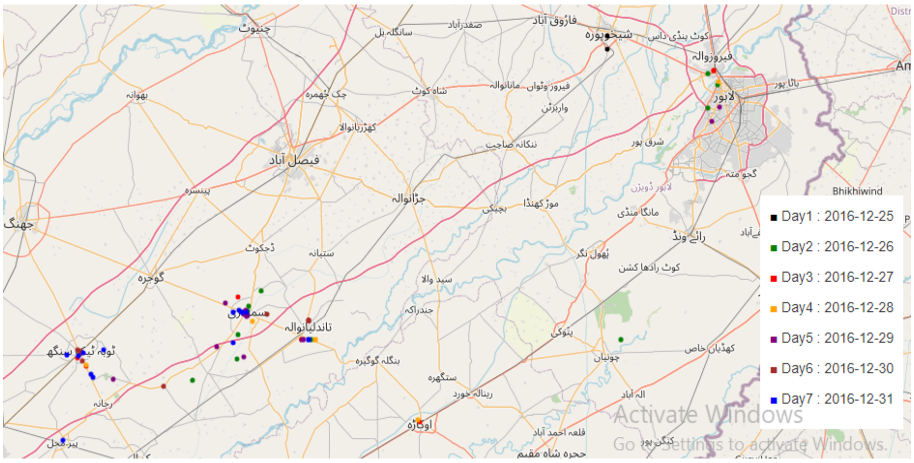


Figure 4.8: Cholera trajectory of one week from 25 December till 31 December,where each color present individual day

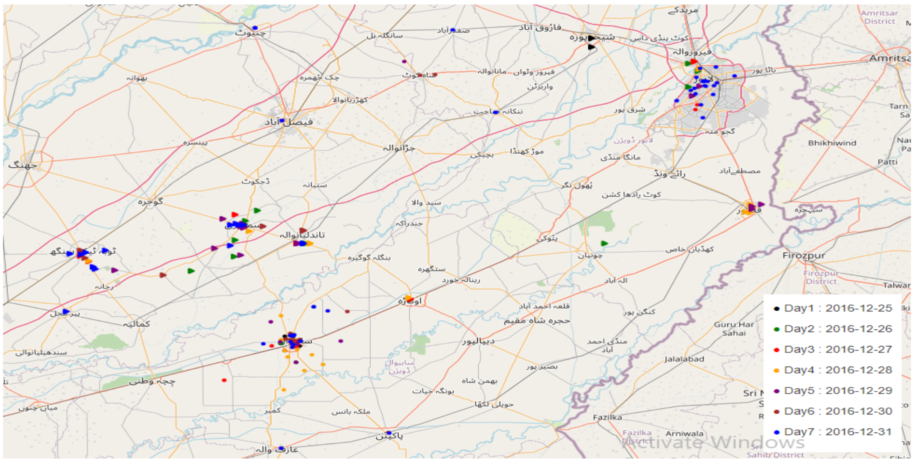


Figure 4.9: Combined SARS and Cholera trajectory of one week.Circle is presenting SARS trajectory while triangle is presenting cholera trajectory

4.0.2 Transformer Results

Figure 4.10 shows comparison of actual and predicted dates.Figure 4.11 shows actual and predicted latitude and longitude, while figure 4.12 shows confusion matrix of disease feature predicted by transformer model.

The below figure 4.13 and 4.14 shows separate one week trajectory for SARS and cholera disease, however figure 4.15 is a combined trajectory for both the diseases.

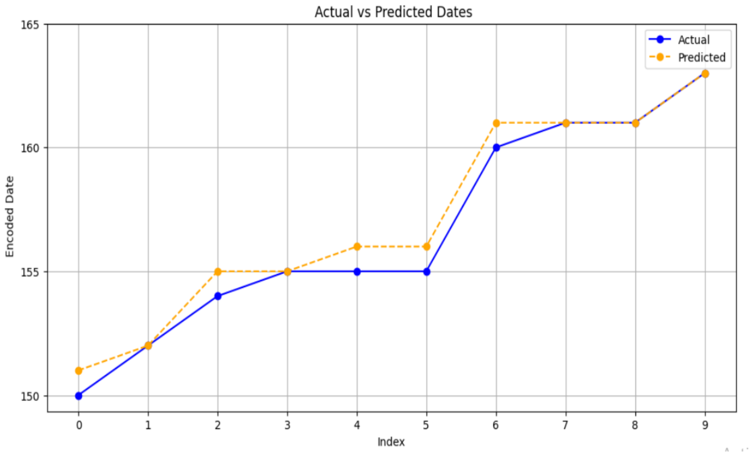


Figure 4.10: Line plot of actual and predicted Dates

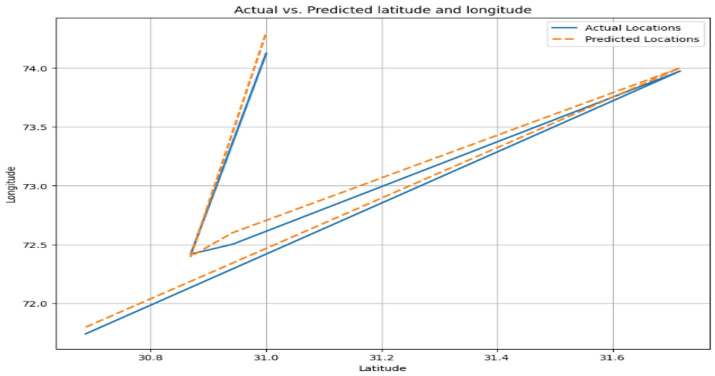


Figure 4.11: Line plot of actual and predicted latitude and longitude

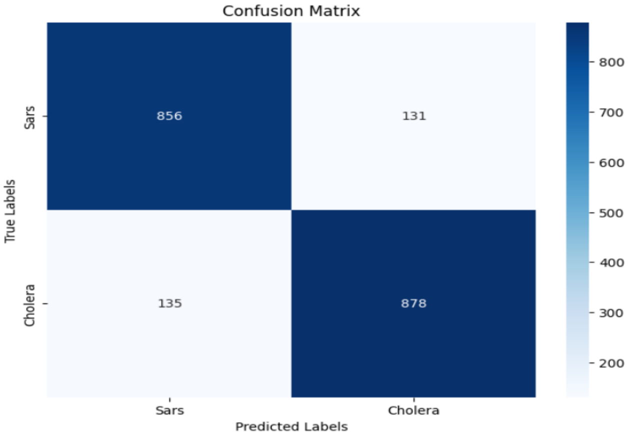


Figure 4.12: Confusion Matrix of Disease predicted by transformer model

The below figure 4.13 and 4.14 shows separate one week trajectory for

SARS and cholera disease, however figure 4.15 is a combined trajectory for both the diseases.

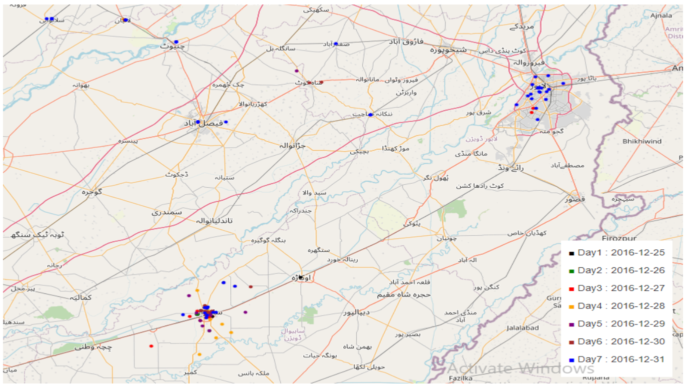


Figure 4.13: SARS trajectory of one week from 25 December till 31 December,where each color present individual day

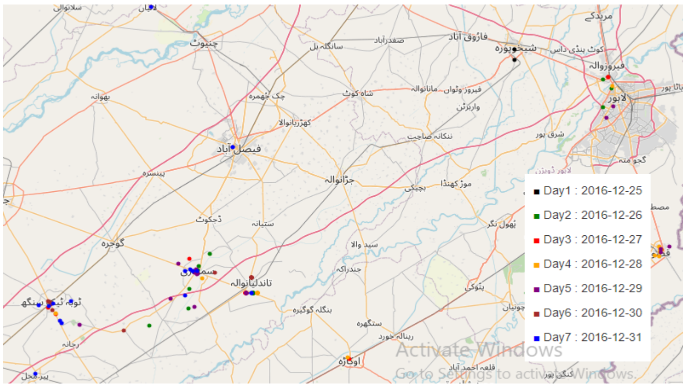


Figure 4.14: Cholera trajectory of one week from 25 December till 31 December

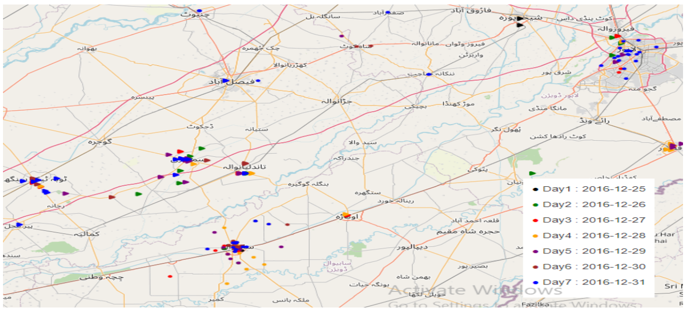


Figure 4.15: Combined SARS and Cholera trajectory of one week.Circle is presenting SARS trajectory while triangle is presenting cholera trajectory

4.0.3 Model Performance

MAE(Mean Absolute Error) is used to measure the error rate of predicted dates, latitude and longitude features.MAE scores were compared for both LSTM and transformer model to check which model performs better. The below table gives a comparison of results for LSTM and transformer model.

Table 4.2: Error comparison of LSTM and Transformer model

Model	MAE of Date	MAE of latitude	MAE of Longitude
LSTM	1.7412	1.1914	.8761
Transformer	1.4675	.7415	.541

Transformer model performed better for predicting dates and disease locations(latitude and longitude). We are also predicting which disease occurred on which location and day, so accuracy, precision, recall and F1 scores are also calculated to measure the models performance.

Table 4.3: Model Performance Metrics Comparison

Model	Accuracy %	Precision	Recall	F1 Score
LSTM	71%	.844	.859	.850
Transformer	76%	.867	.863	.860

The table 4.3 shows LSTM and transformer accuracy,precision, recall,F1 score for predicting disease category(SARS or Cholera). This table presents a comparative analysis of the performance metrics for two models, LSTM and Transformer, in the context of predicting SARS and Cholera diseases.Accuracy metric represents the percentage of correct predictions made by each model. The LSTM model achieved an accuracy of 71%, while the Transformer model outperformed it with an accuracy of 76%. A higher accuracy indicates a better overall predictive performance. Precision measures the model's ability to make accurate positive predictions. For both disease categories (SARS and Cholera), the Transformer model exhibited a slightly higher precision of 0.867 compared to the LSTM model's 0.844. This suggests that the Transformer model had a slightly lower rate of false positives.Recall, also known as sensitivity or true positive rate, assesses the model's capability to correctly identify positive cases. Both models achieved high recall scores, with the LSTM model at 0.859 and the Transformer model at 0.863. A higher recall implies that the models were effective in capturing a significant portion of the actual positive cases.The F1 score is the harmonic mean of precision and recall and provides a balanced measure of a model's performance. The Transformer model demonstrated an F1 score of 0.860, while the LSTM model achieved an F1 score of 0.850. This indicates that the Transformer model maintained a better balance between precision and recall for the given task.

In summary, when predicting SARS and Cholera diseases, the Transformer model demonstrated good performance in terms of accuracy, precision, recall, and F1 score compared to the LSTM model. These metrics collectively suggest that the Transformer model is more effective choice for this predictive task.

CHAPTER 5

CONCLUSION & FUTURE WORK

The potential of deep learning models, particularly LSTM (Long Short-Term Memory) and the transformer, are explores in this work to capture the complex network of SARS and Cholera disease trajectories across several cities of Pakistan. The size of the training data set holds the key to improving disease prediction accuracy. If training is performed on large data set,these models develop the ability to recognize complex spatio-temporal patterns that reflect the spread of disease efficiently.Comparing the performance of LSTM and transformer models reveals that transformers exhibit superior results not only in predicting dates and locations but also in identifying specific disease types, including cholera and SARS. This advantage can be attributed to the multi-head attention mechanism employed by transformers, which is especially well-suited for handling complex time series spatio-temporal data. Unlike LSTM's, transformers excel in capturing both temporal and spatial patterns, making them better equipped to handle the complicated relationships present in such data. By effectively considering the entire context of the input sequence simultaneously, transformers can better capture the complex dependencies and patterns embedded within the data, resulting in enhanced predictive capabilities and improved overall performance.

Enhancing the predictive capability of these models can be achieved through the incorporation of supplementary features. Exploring additional elements, such as population density, local context, and migration patterns, in conjunction with geographical coordinates and disease data, might significantly enhance the models' performance. By exploring deeper into these factors, researchers can gain a better understanding of the dynamics underlying disease outbreaks.To extend the models' applicability to various regions, it is crucial to design them to accommodate multiple locations. This method helps us find out what causes diseases to happen in different areas. Then, we can create plans to help prevent these diseases in specific places.

Furthermore, an encouraging path for more research involves creating combined models that effectively blend the advantages of LSTM and transformer designs. By harnessing the complementary advantages of these two frameworks, researchers can potentially unlock more robust forecasting and prediction capabilities, yielding superior outcomes in disease surveillance and management.

Furthermore, the broader implications of this research extend beyond its immediate geographic focus in Pakistan. The insights and methodologies derived from this study can be applied to diverse global contexts with similar challenges in disease prediction and management. Such cross-contextual applications can catalyze a broader understanding of epidemic dynamics, thus fostering stronger preparedness and response strategies on an international scale.

By broadening the scope of investigation and leveraging interdisciplinary insights, this research can contribute to a comprehensive and holistic understanding of epidemic dynamics and global health crises. As a result, it has the potential to inform evidence-based policy making, strengthen public health interventions, and global healthcare infrastructure against the threat of infectious diseases.

REFERENCES

[1] S. Russell, Artificial Intelligence: A Modern Approach, eBook, Global Edition, Pearson Education, Limited, 2016.

[2] S. A. Bini, Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care?, The Journal of arthroplasty 33 (8) (2018) 2358–2361.

[3] Y. Bengio, Y. Lecun, G. Hinton, Deep learning for ai, Communications of the ACM 64 (7) (2021) 58–65.

[4] G. Atluri, A. Karpatne, V. Kumar, Spatio-temporal data mining: A survey of problems and methods, ACM Computing Surveys (CSUR) 51 (4) (2018) 1–41.

[5] Y. Zheng, L. Zhang, X. Xie, W.-Y. Ma, Mining interesting locations and travel sequences from gps trajectories, in: Proceedings of the 18th international conference on World wide web, 2009, pp. 791–800.

[6] Y. Zhou, H. Wu, H. Cheng, K. Qi, K. Hu, C. Kang, J. Zheng, Social graph convolutional lstm for pedestrian trajectory prediction, IET Intelligent Transport Systems 15 (3) (2021) 396–405.

[7] S. Pellegrini, A. Ess, K. Schindler, L. Van Gool, You'll never walk alone: Modeling social behavior for multi-target tracking, in: 2009 IEEE 12th international conference on computer vision, IEEE, 2009, pp. 261–268.

[8] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, S. Savarese, Sophie: An attentive gan for predicting paths compliant to social and physical constraints, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 1349–1358.

[9] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE transactions on neural networks 20 (1) (2008) 61–80.

[10] P. Trautman, A. Krause, Unfreezing the robot: Navigation in dense, interacting crowds, in: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2010, pp. 797–803.

[11] H. Su, J. Zhu, Y. Dong, B. Zhang, Forecast the plausible paths in crowd scenes., in: IJCAI, Vol. 1, 2017, p. 2.

[12] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, S. Savarese, Social lstm: Human trajectory prediction in crowded spaces, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 961–971.

[13] D. Li, G. Sun, S. Miao, Y. Gu, Y. Zhang, S. He, A short-term electric load forecast method based on improved sequence-to-sequence gru with adaptive temporal dependence, International Journal of Electrical Power & Energy Systems 137 (2022) 107627.

[14] J. Bruna, W. Zaremba, A. Szlam, Y. Lecun, Spectral networks and locally connected networks on graphs international conference on learning representations (iclr2014), CBLS, April (2014).

[15] C. Burstedde, K. Klauck, A. Schadschneider, J. Zittartz, Simulation of pedestrian dynamics using a two-dimensional cellular automaton, Physica A: Statistical Mechanics and its Applications 295 (3-4) (2001) 507–525.

[16] A. Lerner, Y. Chrysanthou, D. Lischinski, Crowds by example, in: Computer graphics forum, Vol. 26, Wiley Online Library, 2007, pp. 655–664.

[17] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, G. Theraulaz, The walking behaviour of pedestrian social groups and its impact on crowd dynamics, PloS one 5 (4) (2010) e10047.

[18] C. Li, H. Zhang, Z. Wang, Y. Wu, F. Yang, Spatial-temporal attention mechanism and graph convolutional networks for destination prediction, Frontiers in Neurorobotics 16 (2022) 925210.

[19] P. O. Ugwoke, F. S. Bakpo, C. N. Udanor, M. C. Okoronkwo, A framework for monitoring movements of pandemic disease patients based on gps trajectory datasets, Wireless Networks (2022) 1–28.

[20] Y.-S. Niu, W. Ding, J. Hu, W. Xu, S. Canu, Spatio-temporal neural network for fitting and forecasting covid-19, arXiv preprint arXiv:2103.11860 (2021).

[21] C. Fritz, E. Dorigatti, D. Rügamer, Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly covid-19 cases in germany, Scientific Reports 12 (1) (2022) 3930.

[22] S. Deng, S. Wang, H. Rangwala, L. Wang, Y. Ning, Cola-gnn: Cross-location attention based graph neural networks for long-term ili prediction, in: Proceedings of the 29th ACM international conference on information & knowledge management, 2020, pp. 245–254.

[23] A. Kapoor, X. Ben, L. Liu, B. Perozzi, M. Barnes, M. Blais, S. O'Banion, Examining covid-19 forecasting using spatio-temporal graph neural networks, arXiv preprint arXiv:2007.03113 (2020).

[24] L. Li, Y. Jiang, B. Huang, Long-term prediction for temporal propagation of seasonal influenza using transformer-based model, Journal of biomedical informatics 122 (2021) 103894.

[25] R. C. Staudemeyer, E. R. Morris, Understanding lstm–a tutorial into long short-term memory recurrent neural networks, arXiv preprint arXiv:1909.09586 (2019).

[26] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[27] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473 (2014).

[28] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).