HealthAssured: Unmasking Misinformation in Healthcare Domain

Submitted in partial fulfilment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

by

Shreyash Divekar, 20102103

Shreyash Gupta, 20102124

Pratham Jain, 20102072

Guide

Prof. Bharti Khemani

Department of Computer Engineering

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE
University of Mumbai

2023-2024

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

CERTIFICATE

This is to certify that the project entitled "**Unmasking Misinformation in Healthcare Domain**" is a bonafide work of **Shreyash Divekar (20102103), Shreyash Gupta (20102124), Pratham Jain (20102072)** submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Engineering.**

Guide	Project Coordinator
Prof. Bharti Khemani	Prof. Rushikesh R. Nikam
Head of Department	Principal
Dr. Sachin H. Malave	Dr. Uttam D. Kolekar

Date:

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

Project Report Approval for B.E.

This project report entitled **"Unmasking Misinformation in Healthcare Domain"** by **Shreyash Divekar, Shreyash Gupta, Pratham Jain** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering**, **2023-2024**.

Examiner Name	Signature
1. ____________________	____________________
2. ____________________	____________________

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words andwhere others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly citedor from whom proper permission has not been taken when needed.

Shreyash Divekar - 20102103

Shreyash Gupta - 20102124

Pratham Jain - 20102072

Date:

Abstract

HealthAssured is an innovative deep learning project designed to address the critical need for accuracy and reliability in the realm of health information. In an era characterised by the rapid proliferation of health-related data and the increasing importance of informed healthcare decision-making, ensuring the veracity of such information is paramount. During this covid pandemic it is clearer than ever how much health misinformation effects. It is much easier now to publish health related articles online without validation, these articles are shared across social media contributing to the spread of health misleading news. In recent research papers, many useful health misinformation detection models use BERT (Bidirectional Encoder Representations from Transformers) which is pretrained on unlabeled data extracted from English Wikipedia and book corpus and are mostly dealt with health misinformation on social media. Therefore, RNN based model that makes use of domain specific word embeddings is proposed for detection of health misinformation specifically in news which is less explored and a dataset combining existing FakeHealth dataset and custom dataset that contains health articles scraped from news fact checking website Snopes.com. Classification results exhibits that the proposed model provides weighted accuracy of 83%.

Keywords: ***Misleading Health Information, RNN-Based Model, Detection, Health Data, Online Health Articles*.**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVATION

ML	Machine Learning
DL	Deep learning
RNN	Recurrent Neural Network
AI	Artificial intelligence
SVM	Support vector Machine
NLTK	Natural Language Toolkit
BERT	Bidirectional-Encoder Representations From Transformers
LSTM	Long short-term memory
RNTN	Recursive-Neural-Tensor Network

CHAPTER 1

Introduction

In today's information-driven world, the accuracy and reliability of health-related data have never been more critical. The inundation of health information, both online and offline, poses a considerable challenge to individuals, healthcare professionals, and organisations. It is against this backdrop that HealthAssured emerged as a pioneering deep learning project with the primary mission of addressing the profound problem of information accuracy in the health domain.

Defining HealthAssured:

HealthAssured is a deep learning project that harnesses the power of neural networks to assess the veracity of health information across various data sources. It is rooted in the conviction that accurate information is the cornerstone of effective healthcare decision-making and improved patient outcomes. By training on diverse datasets that encompass electronic health records, patient-reported data, medical literature, and more, HealthAssured endeavours to classify information as either correct or incorrect, thereby contributing to the trustworthiness of health data. In order to verify the information deep learning models can used, a paper published in the International Journal on Semantic Web and Information Systems[3] explored the enhancement of cyber security through the fusion of Deep Neural Network (DNN) with traditional machine learning methods, which achieved an accuracy of 96.4%.

The proliferation of health information, often riddled with inaccuracies and misinformation, has significant implications for public health and individual well-being. Patients, physicians, researchers, and policymakers rely heavily on accurate health information to make informed decisions. However, the prevalence of misleading data, whether in electronic health records, online health forums, or research publications, can lead to suboptimal healthcare outcomes, unnecessary anxiety, and even health risks. HealthAssured aims to combat this issue by utilizing cutting-edge deep learning techniques to discern the correctness of health-related information. For instance a paper published in IEEE Access explores the application of the Bidirectional Encoder Representation from Transformers (BERT) model, specifically RoBERT, for sentiment classification in news media across four countries. With an impressive accuracy of 90%, the study highlights the effectiveness of the RoBERT model in discerning sentiment nuances in diverse cultural contexts[9].

The primary aim of HealthAssured is to provide a robust solution for verifying the accuracy of health information, benefiting a wide array of stakeholders. This project strives to:

1. Enhance Data Quality:By automatically detecting inaccuracies in health records and other data sources, HealthAssured assists in maintaining high-quality health information.

2. Mitigate Misinformation:HealthAssured serves as a safeguard against the spread of health-related misinformation, promoting responsible health communication.

3. Support Healthcare Professionals:It aids healthcare providers in making well-informed decisions, be it in patient care, research, or telemedicine consultations.

4. Improve Patient Outcomes:By ensuring the accuracy of medication information, diagnoses, and treatment plans, HealthAssured contributes to better patient care and outcomes.

The societal impact of HealthAssured is profound. It bolsters the trustworthiness of health information, thereby empowering individuals to make informed decisions about their health and well-being. Moreover, it assists healthcare professionals in delivering higher-quality care, reducing medical errors, and advancing medical research. In the era of digital health and data-driven decision-making, HealthAssured stands as a beacon of reliability, contributing to the overall improvement of healthcare systems and the well-being of society as a whole.

CHAPTER 2

Literature Survey

Dealing with false information has gained a lot of attention lately, as it has emerged as an alarming problem with increase in volumes of data circulated on internet. This has motivated number of studies to utilize effective technologies aiming to tackel this problem. The successful use of deep learning algorithms in detecting and rectifying disinformation on various digital platform has been demonstrated by prior studies.

For instance, in the domain of Social Media analysis, researchers have leveraged the power of hierarchical attention networks (HAN) combined with multi-layer perceptron (MLP)[1]. HAN, comprising gated recurrent units (GRUs) with attention mechanisms, focuses on context-based features, while MLP emphasizes user-based features. This approach capitalizes on textual and meta data to achieve a notable accuracy of 0.634, with precision, recall, and F1 score metrics of 0.667, 0.556, and 0.607 respectively.

Similarly, in the domain of Cyber Security, the integration of Deep Neural Networks (DNN) with traditional machine learning methods has been explored[3]. By associating DNN with established techniques, the model's hidden layers are empowered to extract more meaningful features from input data. This fusion leads to a heightened accuracy level of 0.964, along with impressive precision and recall scores of 0.93 each, and an F1 score of 0.92.

Another noteworthy advancement comes from the Electrical domain, where Generative Adversarial Networks (GAN) combined with autoencoders (AAE) have been employed for detecting false data feeds[4]. The GAN's discriminator-generator architecture, complemented by AAE's encoding capabilities and network adversarial mechanisms, results in an impressive accuracy of 0.978, with a precision score of 0.95 and a recall score of 0.97.

Furthermore, in the realm of News Media analysis, researchers have utilized Convolutional Neural Networks (CNN) and Naive Bayes classifiers in tandem[5]. Starting with data preprocessing and feature extraction, this approach integrates CNN and Naive Bayes through Q-learning, yielding a commendable accuracy of 0.968. The precision, recall, and F1 score metrics stand at 0.932, 0.921, and 0.92 respectively.

Table 2.1: Literature Survey Table

Journal name Techniques used	**Domain**	**Published Year**	**Summary of Technique**	**Performance Evaluation**
Springer HAN (hierarchical attention network) + MLP (multi-layer perceptron) [1]	Social Media	2022	HAN consists of gated recurrent units (GRUs) with the new attention mechanisms. HAN and MLP uses context-based features and user-based features, respectively. These features are textual or meta data.	Accuracy: 0.634 Precision: 0.667 Recall: 0.556 F1 score: 0.607
Association for Computational Linguistics Relational Graph Convolutional Network [2]	Social Media	2023	PESTO which involves Posts/User and Feature Encoder, which encodes the text and meta features of a post/user into a dense vector, transformer model, RGCN for user-follow network and Fusion Network based on Self-Attention.	Accuracy: 0.915 Precision: 0.912 Recall: 0.921 F1 score: 0.922

International Journal on Semantic Web and Information Systems Deep Neural Network [3]	Cyber Security	2022	Associating DNN with traditional ML methods makes the method of approach more specific as the hidden layers of the model make input features more useful to the model.	Accuracy: 0.964 Precision: 0.93 Recall: 0.93 F1 score: 0.92
IEEE GAN with autoencoder [4]	Electrical	2021	The GAN and AAE algorithm which is responsible for the detection of false data feeds in the system. GAN works with discriminator and generator whereas AAE is for encoding of inputs and adversarial of the network.	Accuracy 0.978 Precision: 0.95 Recall: 0.97

Journal of Soft Computing Paradigm CNN and Naive Bayes [5]	News Media	2022	The process starts with the initial data preprocessing and also feature extraction after which it is worked on with both CNN and Naive Bayes by Q-learning method.	Accuracy: 0.968 Precision: 0.932 Recall: 0.921 F1 score: 0.92
IEEE Journal of biomedical and health informatics OCR(optical character recognition)+ BERT System [6]	Social Media	2020	The text in images was first extracted by an Optical Character Recognition (OCR) algorithm. The OCR results were then processed by a pre-trained BERT system	Accuracy: 0.942 Precision: 0.978 Recall: 0.967 F1 score: 0.973
13th Conference on Knowledge and Smart Technology (KST) XGBoost algorithm [7]	News Media	2021	The author collected samples of 297 reliable and 235 unreliable Experimental results show that XGBoost methods were the most effective.	Accuracy: 0.906 Precision: 0.90 Recall: 0.89 F1 score: 0.89

The 8th IEEE International Conference on E-Health and Bioengineering CNN and neural network. [8]	Health	2020	Recursive Neural Tensor Network (RNTN) and Long short term memory (LTSM) are used.	Accuracy: 0.91 Precision: 0.92 Recall: 0.89
IEEE Access Bidirectional encoder representation Transformer (BERT) [9]	News Media	2021	RoBERT model for sentiment classification and a comparative study on 4 countries.	Accuracy: 0.90
Department of Informatics, Systems, and Communication (DISCo) Long Short-Term Memory Networks (LSTM) & CNN [10]	Health	2022	Long Short- Term Memory Networks is a deep learning, sequential neural network that allows information to persist.	AUC score: 0.84 f-Measure:0.85

Springer GAT (Graph Attention Networks) [11]	Social Media	2024	Graph Attention Networks (GATs) are a type of graph neural network that utilize attention mechanisms to perform node feature aggregation.	Accuracy: 0.84 Macro-F1: 0.92
Emerging Science Journal Bigcn (Bi-Graph Convolutional Network) [12]	Social Media	2024	Bigcn combines top-down with bottom-up tree propagation structure to deal with the propagation and dispersion of rumors.	Accuracy:0.827
Emerging Science Journal LR (Logistics Regression) [13]	News Media	2023	The LR model is used because it provides an equation that is a simple cost function and obtains an equation that categorizes issues between binary or multiple classes.	Accuracy: 0.91 Precision:0.88 Recall:0.93

Journal of Retailing and consumer Service Lstm (Long short-term memory network) [14]	News Media	2024	Long short-term memory network is a recurrent neural network (RNN), aimed at dealing with the vanishing gradient problem present in traditional RNNs.	Accuracy:0.887
International Journal of Intelligent Systems OE-MDL (Optimized Ensemble Machine and Deep Learning) [15]	Social Media	2023	The OE-MDL algorithm enhances detection capabilities by incorporating a series of preprocessing steps: converting text to lowercase,tokenization, eliminating stop words, applying word stemming.	Accuracy: 0.998 Precision: 0.99 Recall: 0.958 F1 score: 0.999

CHAPTER 3

Limitation of Existing System

While existing studies demonstrate promising results in utilizing deep learning and natural language processing techniques to detect and combat health misinformation, several limitations persist within these systems. Here are some key limitations:

- Domain Specificity: Many existing systems are tailored to specific domains within healthcare or social media platforms. This specificity could hinder the applicability of these systems to broader contexts, such as covering a wider range of sources and topics within the healthcare domain.
- Data Quality and Variability: The effectiveness of deep learning models heavily relies on the quality and variability of the training data biases in training data or insufficient representation of certain types of misinformation could impact the performance of the models deployed in various projects.
- Generalization: While some systems achieve high accuracy scores within their specific contexts, their generalizability to new or evolving forms of misinformation remains uncertain.

Addressing these limitations will be essential for the successful development and deployment of projects aimed at combating health misinformation, ensuring their effectiveness in promoting trustworthy information within the healthcare domain.

CHAPTER 4

Problem Statement, Objectives and Scope

4.1 Problem Statement

HealthAssured addresses a multifaceted problem statement within the realm of health information accuracy. This problem statement revolves around the pervasive challenges related to the correctness and reliability of health-related data. In the age of the internet, vast amounts of health-related information are readily available to the public. However, a significant portion of this information is inaccurate, misleading, or even false. Misinformation can range from unproven health remedies shared on social media to incorrect medical advice found on websites. This proliferation of misinformation poses a substantial risk to individuals who may make health decisions based on inaccurate information. Inaccurate or poorly researched medical literature can misguide healthcare professionals and researchers. Articles with flawed methodologies, erroneous conclusions, or incorrect data can lead to misguided clinical decisions, potentially harming patients and hindering medical progress.

4.2 Objectives

The objectives of HealthAssured are centred around its mission to verify the accuracy and reliability of health-related information across various domains. These objectives guide the project's development and implementation, ensuring that it effectively addresses the problem of health information accuracy.

The key objectives of HealthAssured:

1. To build a model capable of detecting misinformation:

Detecting misinformation, particularly in the complex domain of healthcare, is a formidable challenge. This objective is at the core of HealthAssured's mission. To achieve it, the project focuses on developing a robust deep learning model that can effectively identify instances of misinformation within health-related content. This model would need to be trained on a diverse and extensive dataset that encompasses various forms of health information, including textual data from medical articles, patient forums, social media, and potentially visual data like medical images. The model's architecture should be designed to learn patterns, inconsistencies, and characteristics associated with misinformation. Achieving this objective contributes to the prevention of the spread of harmful health-related misinformation, ensuring that individuals receive accurate and reliable health information.

2. To make the dataset more effective by making it more precise and including a variety of data to train upon:

The effectiveness of HealthAssured hinges on the quality and diversity of its training data. This objective emphasises the need to curate a high-quality dataset by improving its precision and expanding its scope. Precision in the dataset can be achieved by meticulously labelling and categorising data points, ensuring that each data entry is accurately annotated as either accurate or inaccurate health information. Moreover, including a wide variety of data sources and formats, such as text, images, and possibly audio, broadens the dataset's applicability and enhances the model's ability to tackle different types of health information. The dataset should encompass various health topics, sources, languages, and styles to make the model more versatile and adaptable.

3. To train the dataset with an appropriate deep learning architecture and achieve the best possible accuracy:

The success of HealthAssured hinges on the choice of a suitable deep learning architecture and the attainment of high accuracy in misinformation detection. This objective underscores the importance of selecting state-of-the-art deep learning techniques tailored to the complexities of health information. HealthAssured aims to optimise the model's architecture, hyperparameters, and training strategies to maximise its accuracy in distinguishing between correct and incorrect health-related information.

4.3 Scope

The scope of HealthAssured is comprehensive, encompassing various aspects of healthcare and health-related information. It is designed to address the critical issue of verifying the correctness and accuracy of health information across different domains.

HealthAssured prioritizes collaboration with fact-checking organizations and platforms specializing in health-related information to combat misinformation effectively. Core emphasis is placed on utilizing fact checkers and web scraping alongside other sources for verifying the correctness and accuracy of health-related data. Their focus spans across various types of information, including clinical data, medical research, treatment information, health advice, and public health data, ensuring thorough assessment and validation. With a primary focus on textual data analysis from diverse sources such as medical records, research articles, online forums, and social media, HealthAssured is well-equipped to address misinformation and ensure data accuracy in the healthcare domain. Serving a wide array of stakeholders including patients, healthcare providers, researchers, and insurance companies, HealthAssured plays a vital role in enhancing trust and reliability within the healthcare ecosystem.

CHAPTER 5

Proposed System

At its core, HealthAssured relies on advanced deep learning and natural language processing (NLP) techniques to analyse and verify textual health-related data. Moreover, HealthAssured collaborates with fact-checking organisations and employs web scraping techniques to gather data from diverse online sources, ensuring a comprehensive approach to misinformation detection. With a focus on precision and scalability, HealthAssured's technological system is built upon a meticulously processed and diverse dataset, allowing it to adapt to a wide range of health information and contexts. At the heart of its operations, HealthAssured integrates advanced deep learning methodologies and natural language processing (NLP) techniques to meticulously scrutinize and authenticate textual data pertinent to the domain of healthcare. By leveraging sophisticated algorithms and models, HealthAssured can discern the nuances within health-related information, thereby enabling the verification of its accuracy and reliability. Moreover, HealthAssured actively collaborates with reputable fact-checking organizations and implements web scraping tools to systematically aggregate data from various online source.

5.1 Architecture

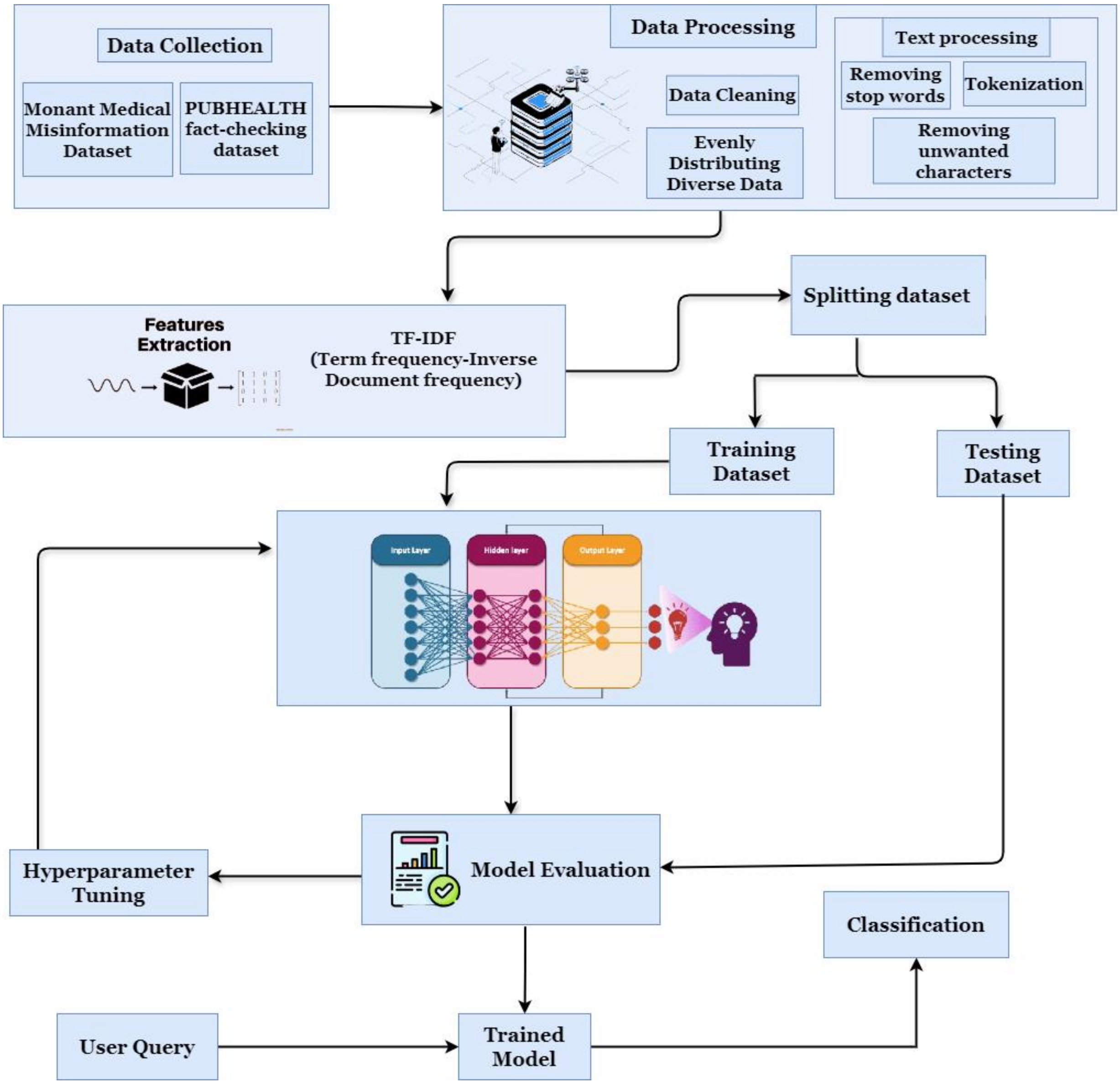


Figure 5.1 – Architecture Diagram

The above architecture diagram of HealthAssured is a well designed system that brings together various components to achieve its core mission of verifying the accuracy of health-related information. While the specifics of the architecture may vary, it typically includes the following key elements:

1. Data Collection: The first step is the collection of health-related data from various sources. This data can include textual information from Fact checkers database which include medical literature, health related articles, social media, and more.
2. Text Preprocessing: Before feature extraction can begin, the textual data is subjected to preprocessing steps to clean and prepare it for analysis. This typically involves:
   - Tokenization: Breaking down the text into individual words or tokens.
   - Lowercasing: Converting all text to lowercase to ensure uniformity.
   - Stopword Removal: Eliminating common words (e.g., "the," "and") that do not carry significant meaning.
   - Lemmatization or Stemming: Reducing words to their base or root forms to simplify analysis (e.g., "running" becomes "run").
3. Feature Engineering: Feature extraction entails the creation of relevant features from the preprocessed text. These features are designed to capture important information and patterns within the data. In HealthAssured, some common techniques used for feature extraction from textual data include:
   - Bag of Words (BoW): BoW represents the frequency of each word in a document, creating a vector for each document. Each word becomes a feature, and the vector's values indicate word frequencies.
   - TF-IDF (Term Frequency-Inverse Document Frequency): TF-IDF is similar to BoW but also considers the importance of a word in the entire dataset. It assigns higher values to words that are specific to a document and less common in the entire dataset.
   - Word Embeddings: Advanced techniques like Word2Vec, GloVe, or fastText can be used to generate word embeddings, representing words as dense, continuous-valued vectors. These embeddings capture semantic relationships between words.
4. Neural network: The neural network in HealthAssured serves as the core engine responsible for evaluating and verifying the accuracy of health-related information.

   Following are sub functionalities performed by neural network:
   - Information Classification: The neural network classifies health data into categories like true and false. It learns to identify patterns and discrepancies in textual and potentially visual data, effectively distinguishing between trustworthy and potentially misleading information.

- Pattern Recognition: Leveraging deep learning, the neural network recognizes intricate patterns and relationships within the data, enabling it to detect subtle signs of misinformation or inaccuracies that might be imperceptible to human observers.

5.2 UML Diagrams

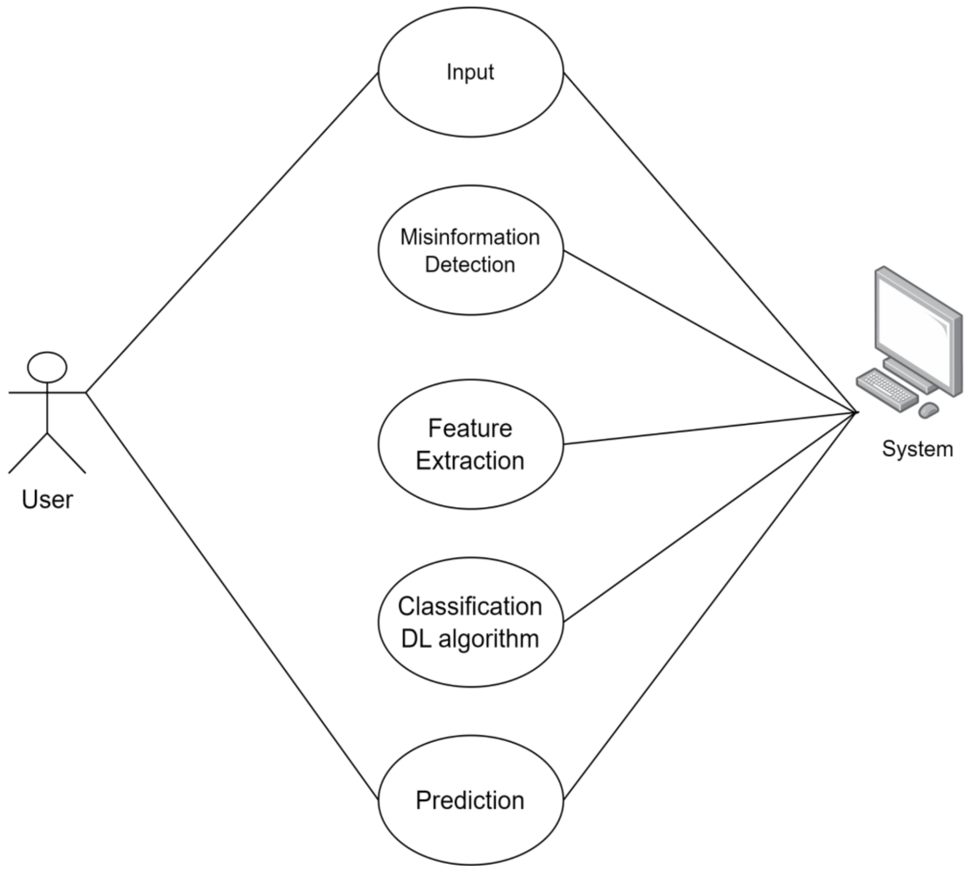


Figure 5.2 Use Case Diagram

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of connections the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses.

The use case diagram visually outlines how different stakeholders interact with HealthAssured to validate and retrieve health information, promoting the project's mission of improving health data accuracy.

CHAPTER 6

Experimental Setup

Algorithm and Process Design

1. Formulating the problem statement

   The core functionality we emphasize is the 'Misinformed Health News Detection' system. This system must possess the capability to identify and flag misleading or false health-related information across various digital platforms. Users should be able to submit potentially misleading news, and the system should employ advanced technologies, including natural language processing and machine learning, to assess the authenticity of these claims. The system should also offer recommendations for verified and trustworthy health resources to assist users in making informed decisions regarding their health.

2. Understanding the framework and Requirement

   Creating a framework for the HealthAssured system involves structuring the project into key components and tasks. Here's a high-level framework for developing the Misinformed health news detection system:

HealthAssured Framework:

(1) Project Initiation:

- Define project objectives and scope.
- Assemble a project team, including data scientists, developers, and domain experts.
- Establish project timelines and milestones.

(2) Data Collection and Integration:

- Identify relevant data sources, such as health news articles, social media, and online forums.
- Develop data collection tools or APIs for real-time data acquisition.
- Implement data integration pipelines to centralize and preprocess incoming data.

(3) Data Preprocessing:

- Clean and normalise text data, removing noise and irrelevant information.
- Tokenization, stop-word removal, and stemming/lemmatization for text data.
- Data quality checks and handling missing values.

(4) Deep Learning Models:

- Develop deep learning models for misleading health news detection.
- Explore natural language processing (NLP) techniques for text analysis.
- Train and fine-tune models on labelled datasets of real and misinformed health news.
- Implement deep learning models (e.g., neural networks) for advanced text analysis.

(5) Evaluation and Improvement:

- Continuously evaluate the system's performance against predefined metrics.
- Use user feedback and data-driven insights to make iterative improvements.

This framework outlines the key steps and components necessary for the development and operation of the HealthAssured misinformation health news detection system. It emphasises the importance of robust data analysis, machine learning,deep learning, scalability, and ongoing maintenance and improvement.

3. Identifying tools/technology to be used

Identifying the right tools and technologies for developing the HealthAssured misinformation health news detection system is crucial for its success. Here are some tools and technologies that you can consider:

(1) Programming Languages:

- Python: Python is widely used for natural language processing and machine learning tasks. Libraries like NLTK, spaCy, and scikit-learn can be employed for text analysis and machine learning.

(2) Machine Learning and AI Frameworks:

- TensorFlow and Keras: These frameworks are excellent for building deep learning models, including neural networks for text analysis.
- PyTorch: An alternative deep learning framework, suitable for NLP tasks and model development.
- Scikit-learn: Useful for traditional machine learning algorithms, such as random forests or support vector machines.

(3) Natural Language Processing (NLP) Libraries:

- NLTK (Natural Language Toolkit): NLTK provides tools for text analysis and natural language understanding.
- spaCy: spaCy is a popular library for NLP tasks, including tokenization, part-of-speech tagging, and named entity recognition.
- Gensim: Gensim is handy for topic modeling and document similarity analysis.

4. Development

The development of the HealthAssured system involves several key stages and steps. Below is a development plan that outlines these stages:

(1) Project Initiation:

- Define project objectives, scope, and requirements.
- Assemble a cross-functional development team with expertise in natural language processing, machine learning, web development, and database management.
- Establish a project timeline and allocate resources.

(2) Data Collection and Integration:

- Identify and source relevant data from health news websites, social media platforms, and other sources.
- Develop data collection scripts or integrate APIs for real-time data acquisition.
- Set up data integration pipelines to centralize and preprocess incoming data.

(3) Data Preprocessing:

- Clean and normalize text data, removing irrelevant information and standardizing formats.
- Perform tokenization, stop-word removal, stemming/lemmatization for text data.
- Implement data quality checks and handle missing values.

(4) Deep Learning Models:

- Develop deep learning models for HealthAssured. Consider models like:
- Natural Language Processing (NLP) models, e.g., LSTM or BERT.
- Traditional machine learning models like Random Forest or SVM.
- Train and fine-tune models on labeled datasets of real and fake health news articles.
- Implement deep learning models (e.g., neural networks) for advanced text analysis.

5. Testing

Perform rigorous testing of the misleading health news detection models.

Conduct system-wide testing to identify and address bugs and issues.

Validate the accuracy and effectiveness of the system through testing and validation with real-world data.

6. Evaluation

The "Evaluation and Improvement" phase of the HealthAssured system is an ongoing process that involves close monitoring of performance, active engagement with users, and the iterative enhancement of the system based on both data-driven insights and user feedback.

7. Details of Hardware and Software

On Windows 64-bit operating systems, in 32-bit or 64-bit mode, the dart runtime requires a minimum of 128MB of memory. The minimum physical RAM is required to run graphically based applications. Running with less memory may cause disk swapping, which has a severe effect on performance. Very large programs may require more RAM for adequate performance. It can run on Microsoft Windows, Solaris OS, Linux or Mac OS. We are running the application on a device with 8 GB ram on an i5 8th Gen Processor on Windows.

CHAPTER 7

Implementation and Result

To achieve the objectives of HealthAssured, deep learning models are being trained on labelled dataset which comprise of broad variety of health related claims and information. The implementation starts with importing all the required libraires for model building and the dataset is loaded to be processed for training and testing the model.

Data Sampling and Label Encoding:

```
num_of_categories = 1000
shuffled = dataframe.reindex(np.random.permutation(dataframe.index))
TRUE = shuffled[shuffled['rating'] == 'TRUE'][:num_of_categories]
FALSE = shuffled[shuffled['rating'] == 'FALSE'][:num_of_categories]
concated = pd.concat([TRUE, FALSE], ignore_index=True)
concated = concated.reindex(np.random.permutation(concated.index))
concated['LABEL'] = 0
concated.loc[concated['rating'] == 'TRUE', 'LABEL'] = 0
concated.loc[concated['rating'] == 'FALSE', 'LABEL'] = 1
```

	claim_id	statements	label
0	15661	$The money#@ the Clinton Foundation took from ...	False
1	9893	$Annual Mammograms$ May Have More FalsePositiv...	False
2	11358	Offers# Prostate $Cancer Patients High Cancer...	False
3	10166	Study Vaccine for Breast Ovarian Cancer Has Po...	True
4	11276	Some# appendicitis$ cases may not require emer...	True

	claim_id	statements	label
0	36252	No public welfare programs existed to help new...	False
1	3639	Former nicotine research monkeys now at primat...	True
2	3948	Flu symptoms sending many to Indianapolisarea ...	True
3	8261	False fears about vaping stopping smokers usin...	True
4	35914	The reason the US documented the most COVID19 ...	False

The procedure starts with preparing a balanced dataset for model building of health-related statements. It first shuffles the dataframe to randomize the data, then selects subsets for 'TRUE' and 'FALSE' labels, each containing a maximum of 1000 samples. These subsets are concatenated into a single dataframe and shuffled again to ensure randomness. A new 'LABEL' column is added, initially set to 0, and then updated based on the 'rating' column values, where 'TRUE' is encoded as 0 and 'FALSE' as 1. This process ensures that the dataset is properly preprocessed with balanced class distribution and encoded labels, ready for subsequent analysis or model training.

Text Tokenization:

```
n_most_common_words = 8000
max_len = 130
tokenizer = Tokenizer(num_words=n_most_common_words, filters='!"#$%&()*+,-
./:;<=>?@[\]^_`{|}~', lower=True)
tokenizer.fit_on_texts(concated['statement'].values)
sequences = tokenizer.texts_to_sequences(concated['statement'].values)
word_index = tokenizer.word_index
```

We start with initializing a Tokenizer object for text processing with specific parameters: num_words set to 8000 to limit the vocabulary size to the most common 8000 words, and filters set to remove punctuation characters from the text. The Tokenizer is then fitted on the concatenated dataframe's 'statement' (which includes the medical claims and information) column to generate a word index, which maps each unique word to a numerical index. Subsequently, the texts_to_sequences method is used to convert the text data into sequences of integers based on the word index, creating a numerical representation of the text. Overall, the procedure tokenizes the textual data, converts it into sequences of integers, and provides insight into the vocabulary size of the dataset.

Model Building:

```
model = Sequential()
model.add(Embedding(n_most_common_words, emb_dim, input_length=X.shape[1]))
model.add(SpatialDropout1D(0.7))
model.add(LSTM(64, dropout=0.7, recurrent_dropout=0.7))
model.add(Dense(2, activation='softmax'))
```

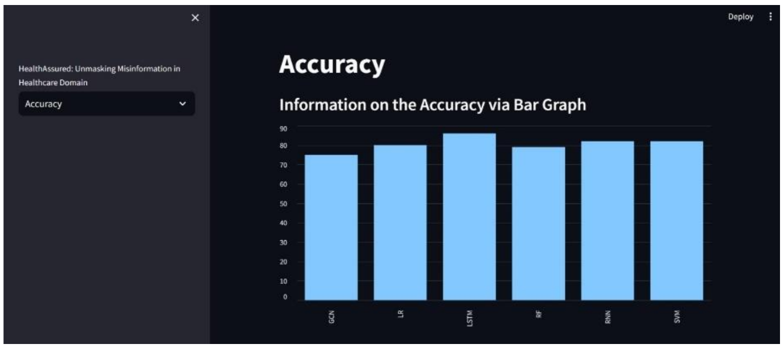
For an instance, while building a recurrent neural network (RNN) model using the Keras Sequential API for text classification tasks. It begins by initializing a Sequential model, which allows stacking layers sequentially. Followed by Embedding layer, converting input sequences into dense vectors. A SpatialDropout1D layer is added for regularization, followed by an LSTM layer with 64 units for learning sequential patterns. Dropout and recurrent dropout are applied in the LSTM layer to prevent overfitting. Finally, a Dense layer with softmax activation is added for output prediction. This model architecture allows for effective learning and clas of health-related statements as either true or false by leveraging embeddings and LSTM units while minimizing overfitting through dropout regularization.

Compiling and Training model for Prediction:

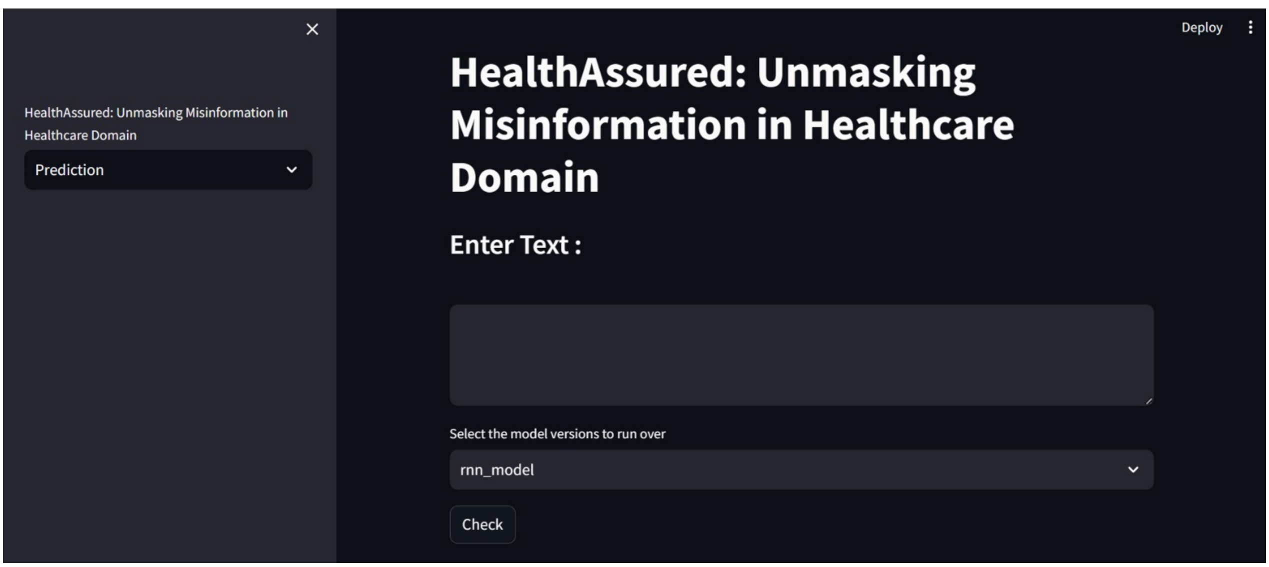
```
# Compile the model
model.compile(optimizer='adam', loss='categorical_crossentropy',
metrics=['acc'])

# Train the model
history = model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size,
validation_split=0.2, callbacks=[EarlyStopping(monitor='val_loss', patience=3,
min_delta=0)])

# Make predictions
txt = ["Does taking apple cider vinegar as a supplement have any real health
benefits?"]
seq = tokenizer.texts_to_sequences(txt)
padded = pad_sequences(seq, maxlen=max_len)
pred = model.predict(padded)
labels = ['TRUE', 'FALSE']
print(labels[np.argmax(pred)])
```

The previously constructed RNN model is compiled using the Adam optimizer, categorical cross-entropy loss function, and accuracy metric for evaluation. Subsequently, the model is trained on the training data (X_train and y_train) with early stopping implemented to prevent overfitting. Early stopping monitors the validation loss, stopping training if it doesn't improve for three consecutive epochs (patience=3). The training process runs for a specified number of epochs (epochs) with a defined batch size. Once the model is trained, predictions are made on a sample text input using the trained model. The input text is tokenized and padded to match the input shape expected by the model. Finally, the predicted class label is printed based on the maximum probability prediction from the model's softmax output layer.

7.1 Implementation



The homepage of HealthAssured gives us an basic overview of the system and tells us about the deep learning models used in the system.



The above snippet is the interface that is built for users to communicate with the model.

The interface includes simple input box which take input from users which is to be tested for veracity. This input is passed on to the model which in turn gives it's real time prediction on the input. This prediction is then displayed on interface below the input box.

7.1.1 Recurrent Neural network

The recurrent neural network (RNN) model implemented in the above code is designed for text classification, specifically for categorising statements into one of three categories: true, false, unproven. The model architecture consists of an embedding layer followed by a spatial dropout layer to prevent overfitting. It then employs a Long Short-Term Memory (LSTM) capturing long-range dependencies in sequential data. The final layer is a dense layer with a softmax activation function, enabling the model to predict the probability of each class.

7.1.2 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network (RNN) designed to effectively handle the complexities of sequential data such as text, time series, and more. With its unique architecture, LSTM can retain information over a long period, making it ideal for processing and classifying textual data into categories like true, false, or unproven. When employed for text classification, LSTM learns to capture the intricate dependencies and patterns within textual data, effectively deciphering the semantic meaning of words and phrases in context. By processing the input text sequentially, LSTM can identify recurring patterns and nuances that contribute to determining the veracity of the information.

7.1.3 Graph Convolutional Networks

Graph Convolutional Networks (GCNs) are a powerful class of deep learning models that operate on graph-structured data. While primarily used for tasks related to graph analysis and node classification, GCNs can also be adapted for text classification tasks that involve understanding relationships and dependencies within textual data. When applied to text classification for determining the veracity of claims or statements, GCNs can leverage the underlying structure of the textual information. By representing the text as a graph, where nodes represent words or phrases and edges represent relationships or connections between them, GCNs can effectively capture the intricate associations and context within the text.

7.1.4 Support Vector Machine

Support Vector Machine (SVM) is a powerful supervised machine learning model used for classification and regression tasks. It operates by finding an optimal hyperplane that best separates different classes or groups of data points. When applied to text classification, SVM can effectively discern patterns and features in textual data to classify it into distinct categories such as true, false, or unproven. By transforming textual data into numerical feature vectors, SVM can analyze the relationships and patterns within the data. It aims to identify the most suitable hyperplane that maximizes the margin between different classes, thereby enabling accurate classification of text into the desired categories

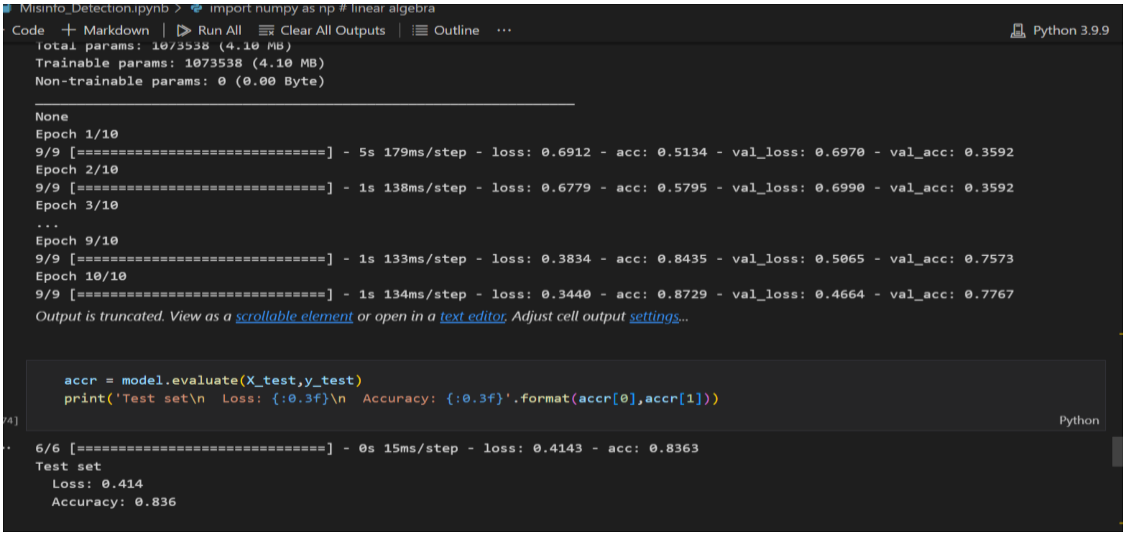
7.1.5 Logistic Regression

Logistic Regression is a fundamental machine learning model primarily used for binary classification tasks. Although it's not as complex as some deep learning models, it remains a powerful tool for various classification problems, including text classification. When employed in the context of classifying textual data into categories such as true, false, or unproven, Logistic Regression can effectively model the relationship between the features extracted from the text and the probability of it belonging to a particular class. By transforming textual data into numerical feature vectors, it computes a weighted sum of the input features and applies a logistic function to predict the probability of the input belonging to a particular category. While it might not capture intricate relationships or complex patterns within the text as effectively as deep learning models

7.1.6 Random Forest

Random Forest is a versatile and robust machine learning algorithm commonly used for both classification and regression tasks. In the context of classifying text into categories such as true, false, or unproven, Random Forest can effectively handle the complexities of textual data and provide accurate predictions. By constructing multiple decision trees during the training process and combining their outputs through a voting mechanism, Random Forest can make robust predictions based on the collective knowledge of multiple trees. Each decision tree is trained on different subsets of the data, thereby reducing overfitting and increasing the model's generalisation capabilities.

7.2 Results



```
Misinfo_Detection.ipynb > import numpy as np # linear algebra
Code + Markdown | Run All  Clear All Outputs | Outline ...                Python 3.9.9
Total params: 1073538 (4.10 MB)
Trainable params: 1073538 (4.10 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
None
Epoch 1/10
9/9 [==============================] - 5s 179ms/step - loss: 0.6912 - acc: 0.5134 - val_loss: 0.6970 - val_acc: 0.3592
Epoch 2/10
9/9 [==============================] - 1s 138ms/step - loss: 0.6779 - acc: 0.5795 - val_loss: 0.6990 - val_acc: 0.3592
Epoch 3/10
...
Epoch 9/10
9/9 [==============================] - 1s 133ms/step - loss: 0.3834 - acc: 0.8435 - val_loss: 0.5065 - val_acc: 0.7573
Epoch 10/10
9/9 [==============================] - 1s 134ms/step - loss: 0.3440 - acc: 0.8729 - val_loss: 0.4664 - val_acc: 0.7767
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

accr = model.evaluate(X_test,y_test)
print('Test set\n  Loss: {:0.3f}\n  Accuracy: {:0.3f}'.format(accr[0],accr[1]))
                                                                      Python
6/6 [==============================] - 0s 15ms/step - loss: 0.4143 - acc: 0.8363
Test set
  Loss: 0.414
  Accuracy: 0.836
```

Accuracy of 83% was achieved using RNN (recurrent neural network) model .

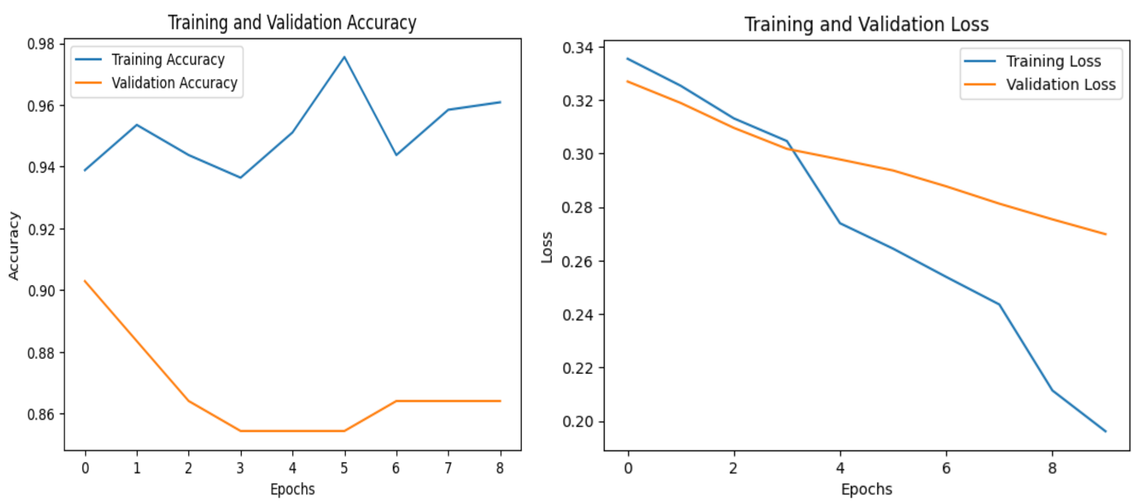


Figure 7.2.1 Training Loss and Accuracy graph for RNN

The RNN model exhibits promising progress with a low loss of 0.419, demonstrating efficient learning and strong accuracy of 0.819, indicating its adeptness in capturing complex data patterns.

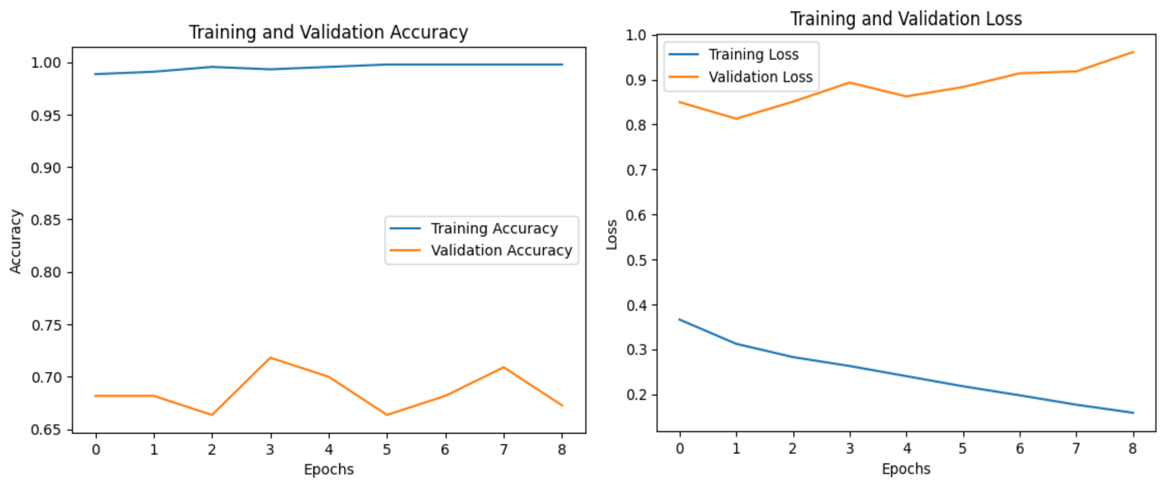


Figure 7.2.2 Training Loss and Accuracy graph for LSTM

The LSTM model shows highly encouraging results with a low training loss of 0.35, showcasing exceptional proficiency in learning complex sequential data patterns, and an impressive training accuracy of 0.86, emphasizing its strong predictive capabilities.

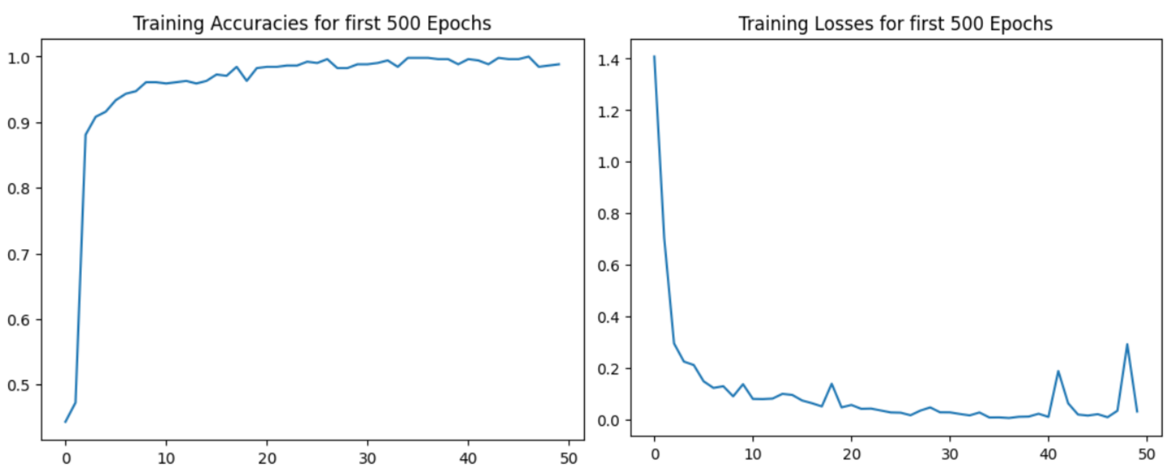


Figure 7.2.3 Training Loss and Accuracy graph for GCN

The GCN model exhibits promising capability with a low training loss of 0.29, effectively minimizing errors in learning complex graph-structured data patterns. Yet, its accuracy at 0.65 suggests some challenges in accurately predicting target variables within the graph.

COMPARATIVE STUDY:

Table 7.2 Comparative study table

MODEL		CLASS	ACCURACY	PRECISION	RECALL	F1 - SCORE	SUPPORT	AUC SCORE
ML	SVM	0	82	0.89	0.90	0.89	69	-
		1		0.90	0.88	0.89	68	
	LOGISTIC REGRESSION	0	80	0.85	0.88	0.87	69	
		1		0.88	0.84	0.86	68	
	RANDOMFO REST	0	79	0.86	0.80	0.83	69	
		1		0.81	0.87	0.84	68	
DL	LSTM	0	86	0.92	0.81	0.86	69	0.86
		1		0.83	0.93	0.88	68	
	RNN	0	82	0.91	0.72	0.80	88	0.82
		1		0.75	0.93	0.83	83	
	GCN	0	75	0.74	0.80	0.77	83	0.75
		1		0.70	0.84	0.70	88	

A thorough examination of various machine learning and deep learning models underscores their distinct capacities in processing complex data structures. The Long Short-Term Memory (LSTM) model demonstrated exceptional performance, recording an impressive accuracy of 86%. This high accuracy underlines its proficiency in capturing and interpreting intricate temporal dependencies within sequential data, making it particularly suitable for tasks involving time-series analysis and natural language processing. In parallel, the Recurrent Neural Network (RNN) achieved a commendable accuracy of 82%, reaffirming its capability to analyse sequential data, although slightly less adept than the LSTM in modelling long-term dependencies. On the other hand, the Graph Convolutional Network (GCN) displayed comparatively lower accuracy at 65%, suggesting the need for further enhancements to enable more effective learning of complex relationships within graph-structured data.

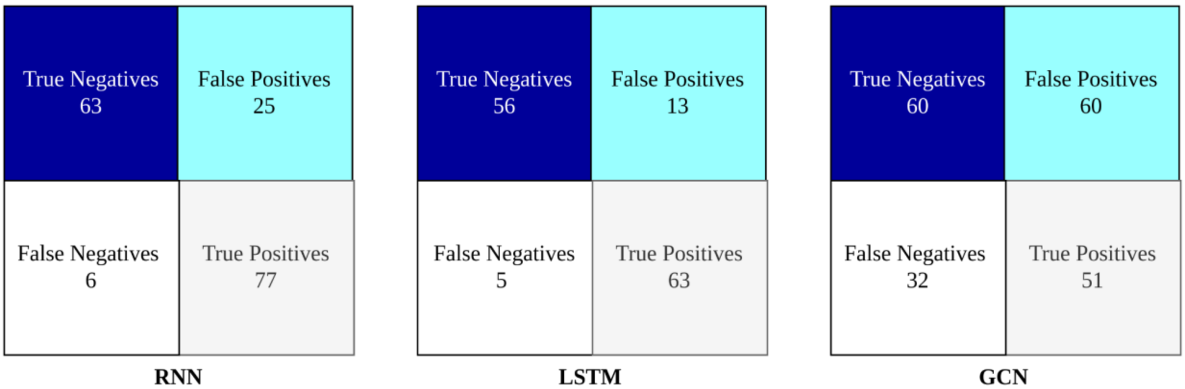


Fig 7.2.4 Confusion Matrices

In deep learning evaluation, TP, FP, TN, and FN comprise the confusion matrix, reflecting correct and incorrect classifications. These metrics enable the calculation of accuracy, precision, recall, and F1 score, crucial for assessing model performance comprehensively. Understanding these parameters aids in fine-tuning models to improve predictive capabilities and optimize overall performance.

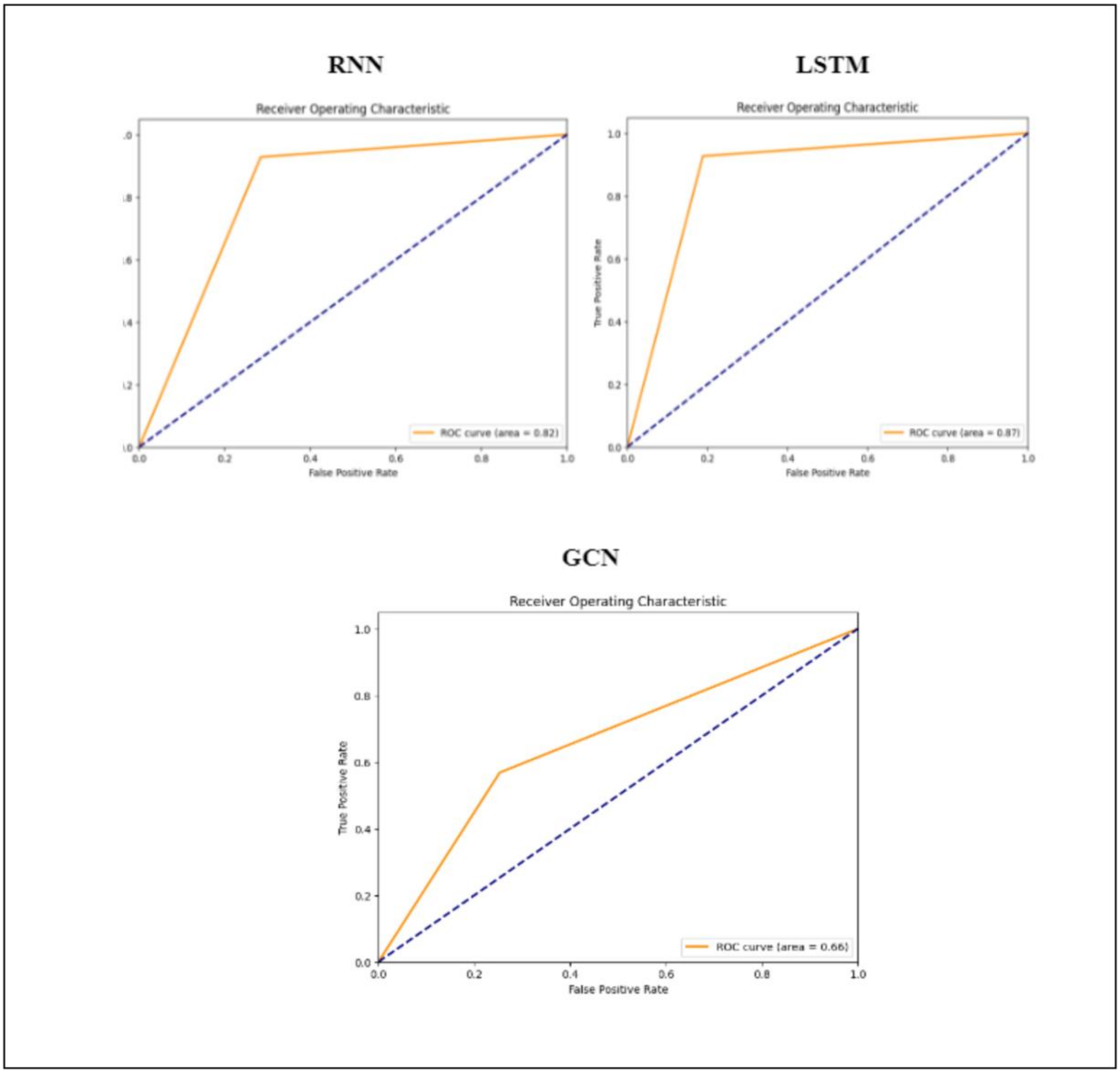


Fig 7.2.5 AUC-ROC curves

The AUC-ROC values reveal the performance of deep learning models. RNN achieves 0.82, showing proficiency in sequential data handling, while LSTM surpasses with 0.86, indicating superior long-term dependency capture, ideal for tasks like time-series forecasting and NLP. GCN obtains 0.66, indicating capability in graph data but with limitations in class discrimination, though it holds promise for applications with interconnected data.

CHAPTER 8

Project Plan

- **Gantt Chart**

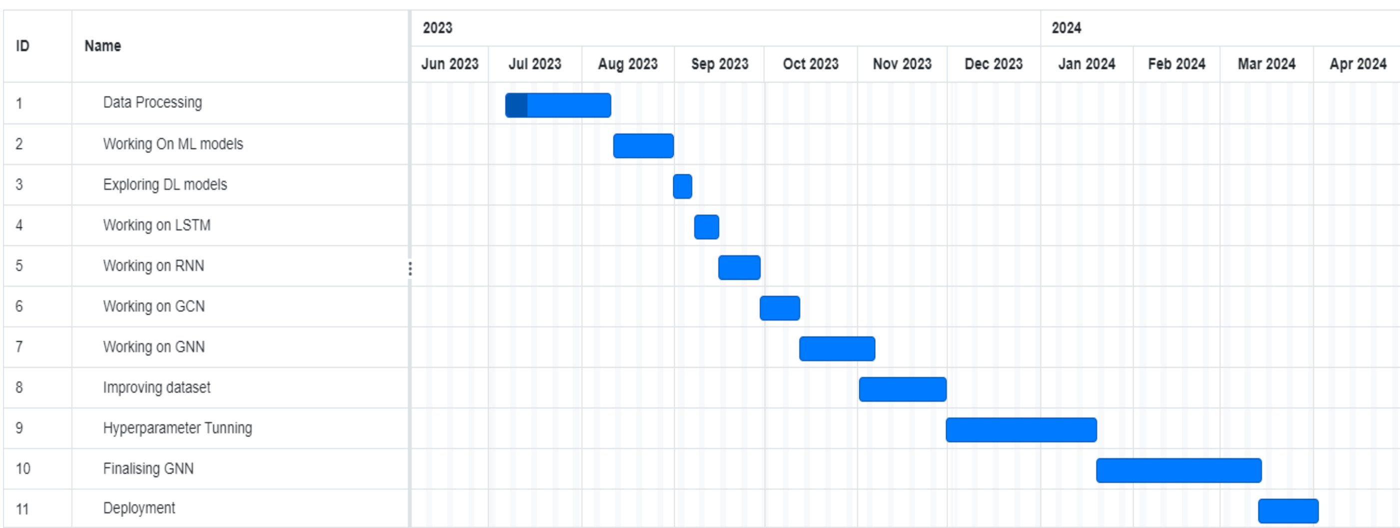


Figure 8.1 – Gantt Chart

The Gantt chart illustrates project tasks with corresponding start and end dates, along with durations for each task, facilitating sequential or concurrent completion to achieve project objectives within the specified timeframe, from the project's start to its conclusion.

CHAPTER 9

Conclusion

HealthAssured has explored diverse machine learning and deep learning algorithms with the purpose of unmasking misinformation in health domain. The project was mainly focused on deep learning algorithm for achieving the objective of misinformation detection, the process of exploring these diverse algorithms involved developing capable deep learning models of RNN, LSTM and GCN. Each explored model excelled in respective areas of misinformation detection. Different methods and approaches towards the same problem statement made a comparative study possible amongst multiple deep learning and machine learning models. The results clearly showcase supremacy of deep learning models over machine learning models.

CHAPTER 10

Future Scope

Enhancing the effectiveness of the model requires a strategic focus on improving the dataset it relies on for training. This involves several key steps aimed at enriching the dataset's diversity, ensuring data quality, refining annotations, exploring data augmentation techniques, and fostering collaborative data-sharing initiatives. Firstly, expanding the dataset's diversity is crucial. By incorporating data from various healthcare sources such as medical journals, reputable websites, forums, and social media platforms, the model gains exposure to a broader range of healthcare-related information. This diverse dataset provides a more comprehensive view of the healthcare landscape, enabling the model to better detect misinformation across different platforms and contexts. Additionally, including data from multiple languages and regions helps address cultural nuances and regional variations in healthcare information, enhancing the model's effectiveness in diverse global settings. Secondly, maintaining data quality is paramount. Implementing rigorous data validation processes ensures the authenticity and reliability of the collected data. Fact-checking procedures and manual verification by domain experts help filter out unreliable or biased information, improving the accuracy of the model's predictions. Continuously updating and refining the dataset with new information and trends in the healthcare domain ensures that the model remains up-to-date and adaptive to evolving misinformation patterns. Thirdly, enhancing annotation efforts is essential for training a robust misinformation detection model. Annotating the dataset with granular labels that capture various types and levels of misinformation, such as false claims, misleading information, and biased opinions, enables the model to distinguish between different degrees of misinformation accurately. Incorporating user feedback mechanisms further refines the model's understanding by validating its predictions and incorporating real-world insights. Additionally, exploring data augmentation techniques can help mitigate data scarcity issues and improve the model's robustness to variations in input data. Techniques such as synthetic data generation and data augmentation through transformations or perturbations of existing data samples can artificially expand the dataset, providing additional training examples for the model. Lastly, fostering collaborative data-sharing initiatives with healthcare institutions, research organizations, and public health agencies can enrich the dataset with high-quality healthcare data and facilitate knowledge exchange on misinformation detection strategies. Collaborations enable access to diverse data resources and domain expertise, strengthening the model's training dataset and ultimately enhancing its performance in detecting healthcare misinformation.

References

[1] Jain, D.K., Kumar, A. & Shrivastava, A. CanarDeep: a hybrid deep neural model with mixed fusion for rumour detection in social data streams. Springer Publications. Neural Comput & Applic 34, 15129–15140 (2022). https://doi.org/10.1007/s00521-021-06743-8

[2] Erxue Min and Sophia Ananiadou. 2023. PESTO: A Post-User Fusion Network for Rumour Detection on Social Media. In Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 1–10, Toronto, Canada. Association for Computational Linguistics.

[3] Meng, Weizhi & Kwok, Lam. (2022). Adaptive False Alarm Filter Using Machine Learning in Intrusion Detection. The 6th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). Shanghai China

[4] Y. Zhang, J. Wang and B. Chen, "Detecting False Data Injection Attacks in Smart Grids: A Semi-Supervised Deep Learning Approach," in IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 623-634, Jan. 2021, doi: 10.1109/TSG.2020.3010510.

[5] Hamdan, Yasir & Ammayappan, Sathesh. “Faultless Decision Making for False Information in Online” 2022, Journal of Soft Computing Paradigm 2(4):226-235 doi:10.36548/jscp.2020.4.004

[6] Z. Wang, Z. Yin and Y. A. Argyris, "Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 2193-2203, June 2021, doi: 10.1109/JBHI.2020.3037027.

[7] Saengkhunthod, Chotipong & Kerdnoonwong, Parischaya & Atchariya Chanwanich, Kanokwan. “Detection of Unreliable Medical Articles on Thai Websites" 2022. 13th International Conference on Knowledge and Smart Technology (KST). doi:10.1109/KST51265.2021.9415756

[8] Y. Parfenenko, A. Verbytska, D. Bychko and V. Shendryk, "Application for Medical Misinformation Detection in Online Forums," 2020 International Conference on e-Health and Bioengineering (EHB), Iasi, Romania, 2020, pp. 1-4, doi: 10.1109/EHB50910.2020.9280120.

[9] P. Ghasiya and K. Okamura, "Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach," in IEEE Access, vol. 9, pp. 36645-36656, 2021, doi:10.1109/ACCESS.2021.3062875.

[10] Di Sotto S, Viviani M. Health Misinformation Detection in the Social Web: An Overview and a Data Science Approach. Int J Environ Res Public Health. 2022 Feb 15;19(4):2173. doi: 10.3390/ijerph19042173.

[11] Lakzaei, Batool, Mostafa Haghir Chehreghani, and Alireza Bagheri. "Disinformation detection using graph neural networks: a survey." Artificial Intelligence Review 57.3 (2024): 52.

[12] Liu, Tianrui, et al. "Rumor Detection with a novel graph neural network approach." arXiv preprint arXiv:2403.16206 (2024).

[13] Fahad, Nafiz, et al. "Stand up against bad intended news: An approach to detect fake news using machine learning." Emerging science journal 7.4 (2023): 1247-1259.

[14] Choudhary, Anshika, and Anuja Arora. "Assessment of bidirectional transformer encoder model and attention based bidirectional LSTM language models for fake news detection." Journal of Retailing and Consumer Services 76 (2024): 103545.

[15] Ganpat, Raut Rahul, and Sonawane Vijay Ramnath. "OE-MDL: Optimized Ensemble Machine and Deep Learning for Fake News Detection." International Journal of Intelligent Systems and Applications in Engineering 12.12s (2024): 60-85.

Publication

The project titled 'Unmasking Misinfornation in Healthcare domain' has been published in IEEE Conference, '5th 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)'

GALGOTIAS UNIVERSITY · विकसित भारत अभियान · IC2PCT · E Cell · INSTITUTION'S INNOVATION COUNCIL GALGOTIAS UNIVERSITY · WIE 25 YEARS · IEEE UP SECTION (INDIA)

5th IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT-2024)

Conference Record #60090

February 9-10, 2024

Certificate of Participation

This is to certify that *Shreyash Vijaykumar Divekar* of A.P Shah Institute of Technology presented his/her research paper titled **"Unmasking Healthcare Misinformation RNN Based Detection"** in the IC2PCT-2024 organized by Galgotias University, Greater Noida, Uttar Pradesh, India on 9th & 10th February, 2024.

Dr. Aanjey Mani Tripathi Chair, IC2PCT-2024	Dr. Avadhesh Kumar PVC & General Chair, IC2PCT-2024	Dr. K. Mallikharjuna Babu VC & Patron, IC2PCT-2024

Unmasking Healthcare Misinformation RNN Based Detection

Prof. Bharti Khemani
Symbiosis International (Deemed University), Pune 412115, India; bharti.khemani.phd2021@sitpune.edu.in

Shreyash Divekar
Department of Computer Engineering
A. P. Shah Institute of Technology
Mumbai, India
shreyashdivekar2002@gmail.com

Pratham Jain
Department of Computer Engineering
A. P. Shah Institute of Technology
Mumbai, India
prathamjain8292@gmail.com

Shreyash Gupta
Department of Computer Engineering
A. P. Shah Institute of Technology
Mumbai, India
shreyashgupta60@gmail.com

Prof. Manasi Chouk
Department of Computer Engineering
A. P. Shah Institute of Technology
Mumbai, India
mvchouk@apsit.edu.in

Dr. Shruti Patil
Symbiosis Institute of Technology Pune Campus, Symbiosis International (Deemed University) (SIU), Lavale, Pune 412115, India.
shruti.patil@sitpune.edu.in

* *Corresponding Authors : Bharti Khemani;bharti.khemani.phd2021@sitpune.edu.in; Shruti Patil; shruti.patil@sitpune.edu.in*

***Abstract*— In today's world, internet has opened up a box full of data to users. In field of healthcare the authenticity of data is very crucial for patient well being and well informed decision making. This research involves the use of neural networks towards the objective of identifying misinformation in healthcare domain. The study emphasizes the importance of precise and diverse data to build a model capable of unmasking misinformation in healthcare domain. The primary goal is to offer a robust method for evaluating the veracity of health information. Through an comparative analysis of machine learning and deep learning models, the research concludes that the RNN deep learning model excels in identifying false healthcare information. The study showcase the superiority of deep learning models over traditional machine learning approaches, underscoring the need for enhancing datasets to enhance the accuracy and relevance of the detection system, due to ever changing landscape of healthcare data.**

Keywords— Misleading Health Information, RNN-Based Model, Detection, Health Data, Online Health Articles

Introduction

This study explores the field of neural networks to ascertain the validity of health information gathered from various sources. In health domain accuracy of data is very decisive for patient well being, fatal consequences can be suffered due to inaccurate data in health domain. The research works on this issue by drawing on concepts from a range of datasets classified by trusted fact checkers. Abundance of health data, both digital and analog, has proven to be double edged swords which can positively help patients to be well informed or cause fatal loses due to inaccuracies.

In order to overcome this issue, this research evaluates health-related data using newly developed techniques. The research aims to achieve the following objectives:

• Improve Data Quality: By automatically identifying mistakes in medical records and other data sources, this research contributes to the maintenance of correct health information.

• Counter Misinformation: This study serves as a check on the spread of misleading health information by promoting the appropriate dissemination of health-related facts.

• Help Medical Professionals: Healthcare practitioners can use it to make more educated judgments whether it comes to patient care, research initiatives, or telemedicine consultations.

• Enhance Patient results: By ensuring the accuracy of prescription information, diagnoses, and treatment plans, this research enhances patient care and overall result.

Related Work

Dealing with false information has gained a lot of attention lately, as it has emerged as an alarming problem with increase in volumes of data circulated on internet. This has motivated number of studies to utilize effective technologies aiming to tackel this problem..

Table.1. Literature Survey

Techniques used	Domain	Year	Summary of Technique	Performance Metrics
Springer HAN (hierarchical attention network) + MLP (multi-layer perceptron) [4]	Social Media	2022	HAN consists of gated recurrent units (GRUs) with attention mechanisms.HAN and MLP uses context-based features and user-based features, respectively. These features are textual or meta data.	Accuracy: 0.634 Precision: 0.667 Recall: 0.556 F1 score: 0.607
Association for Computational Linguistics Relational Graph Convolutional Network [15]	Social Media	2023	PESTO involves Posts/User Feature Encoder, which encodes the text and meta features of a post/user into a dense vector, transformer model, RGCN for user-follow network and Fusion Network based on Self-Attention.	Accuracy: 0.915 Precision: 0.912 Recall: 0.921 F1 score: 0.922
International Journal on Semantic Web and Information Systems Deep Neural Network [16]	Cyber Security	2022	Associating DNN with traditional ML methods makes the method of approach more specific as the hidden layers of the model make input features more useful to the model.	Accuracy: 0.964 Precision: 0.93 Recall: 0.93 F1 score: 0.92
IEEE GAN with autoencoder [17]	Electrical	2021	The GAN and AAE algorithm is responsible for detecting false data feeds in the system. GAN works with discriminator and generator	Accuracy 0.978 Precision: 0.95 Recall: 0.97

The successful use of deep learning algorithms in detecting and rectifying disinformation on various digital platform has been demonstrated by prior studies. Instance of these Recursive Neural Tensor Network (RNTN) and Long Short-Term Memory (LSTM) network are two methods. Effective deep learning and natural language processing (NLP) techniques are used by system to assess and verify textual health-related data. It works with fact-checking organization and uses web scraping techniques to collect data from server online source to guarantee a comprehensive approach to misinformation identification. A 2020 research report [2] looks at several important topics, Including finding medical misinformation in online forms.

Two well-Known algorithms that can handle sequential data and extract pattern from it are Long-Short-Term Memory and Recursive Neural Tensor Networks, in their research, they employ both approaches. Their research is important because of the high accuracy score 0.91, which shows that their approach can successfully identify and report medical misinformation in the complex web of online forums.

In-depth information about identifying health misinformation on social media is provided by another study [3]. Long short-term memory networks, a kind of neural network well-known for its capacity to absorb and retain information, are the focus of this study. The study's AUC (Area under the curve) score of 0.84, which shows that LSTM network is efficient in stopping the spread of false information about health on various social media platforms, makes it remarkable.

A recent study from 2022 introduces "CanarDeep,"[4] a state-of-the-art hybrid deep neural network designed exclusively to identify rumors in real-time social data streams. Their integration of attention processes with a Hierarchical Attention Network (HAN) composed of gated recurrent units (GRUs) is a fundamental aspect of their system. By combining these several factors, CanarDeep can effectively recognize and classify rumors, marking a significant advancement int the field of social data analysis rumor detecting approaches.

PROPOSED SYSTEM

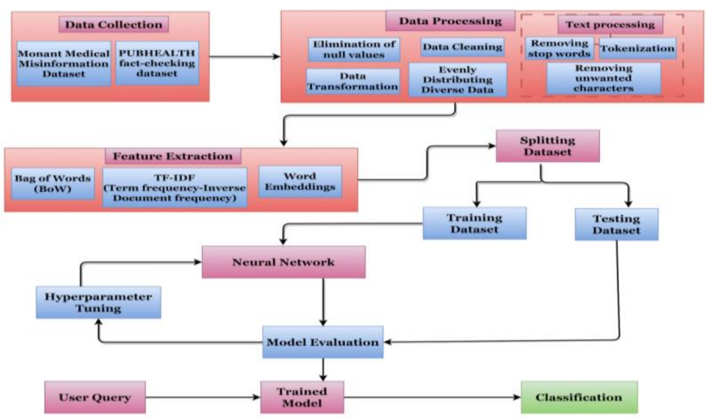


Fig. 1. Architecture Diagram

Natural language processing (NLP) techniques and learning methodologies are the foundation of its operations; these are used to deeply analyze and validate textual information relevant to the healthcare industry. It is possible to confirm the accuracy and dependability of the data by using complex models and algorithms that may identify subtleties in health-related data. Additionally, it actively works with respectable fact-checking groups and uses web scraping technologies to methodically gather data from numerous online sources. This combination of method guarantees a comprehensive and in-depth examination of health-related content, simplifying the process of identifying errors and misleading information that are frequently encountered in the world of digital media.

Our system is a well-designed system that incorporates several. components to achieve its main objective of verifying the accuracy of health-related data. The architecture diagram of our system is displayed above. The main components of architecture often include the following; however, specifics can vary:

3.1 Data Collection: The initial step involves compiling health-related data from various sources, such as the Monant Medical Misinformation Dataset [24]. Textual data from the fact checker database may be included in this data; this dataset include social media, health-related articles, medical literature, and more.

3.2 Text Preprocessing: Before feature extraction can begin, the textual data must be cleaned up and prepared for analysis through preprocessing steps. Usually, this comprises:

• The method of tokenizing a text involves breaking it up into distinct words or unit, such as TF-IDF (Term Frequency-Inverse Document Frequency): Bow-like TF-IDF also considers the importance of a word over the whole dataset. Words that are exclusive to a document and less common throughout the collection receive higher values.

• Word Embeddings: This is an NLP component which involves the preprocessing of text. It is performed to capture semantic information of the data by representing them in lower dimensional space

• Lowercasing: In order to maintain consistency we convert all text into lower cases. This avoids unnecessary discrimination of words

• Stopword Removal: Stopwords are the terms that do not have any significant meaning to convey. They are used for grammatical purpose during sentence composition

• Lemmatization or stemming: It is the process of reducing a word to its root form. The purpose of doing so is to reduce amount of words to process

3.3 Feature Engineering: Other than predefined features this process identifies more of them. A model gives better and faster performance due to this process as it provides data points to use for the task. The process include methods like:

• Bag of Words (BoW): It is based on putting words into categories based on its frequency in the corpus. With this technique it is easier to deal with text and handle large corpus

3.4 Neural network:

Neural networks are designed to process information like human brain. Each layer of the neural network works on improving the output produced. Every layer is closely connected to the other such that they process in sequential patterns.

The process involves different encoders which collectively assist in forming the resultant prediction. In order to simplify the task it is divided into small nodes where it is processed individually. After which each outcomes are combined to form a collective output which fulfills the complex task

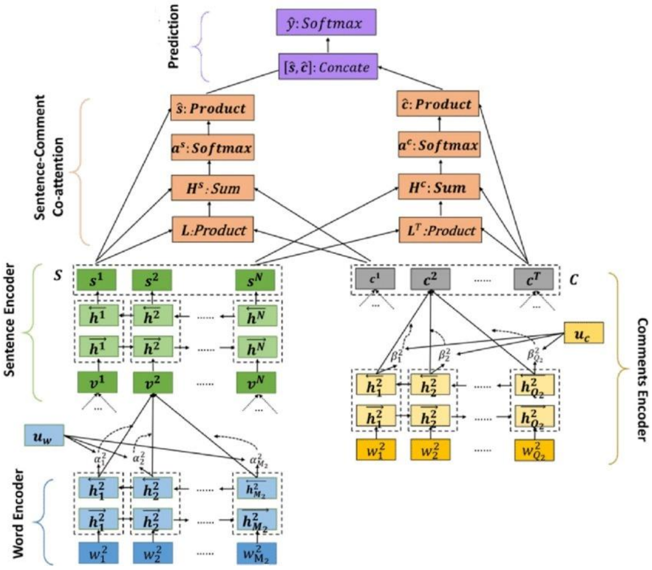


Fig. 2. Working of RNN model

The above diagram represents steps in which our model process the data and work on producing the output.

METHODOLOGY

Our research was inclusive of a variety of algorithms which were designed to process the input and complete the objective of identifying misinformation. While studying and implementing each algorithm we noticed that each of them had respective advantages and disadvantages. The depth of every algorithm was different in terms of complexity of it;s architecture. Let us discuss every explored algorithm in detail.

4.1 Long Short-Term Memory

LSTM is a deep learning algorithm having multiple layers of neural networks, it has a very complex and detailed architecture which is capable of processing sequential data, this capability of LSTM enables it to retain information. We designed the architecture of LSTM with a view of completing the objective of misinformation detection. Processing textual data is effectively done by sequential data processing of LSTM.

4.2 Recurrent Neural network

RNN is an simplified version of LSTM where RNN is designed to perform basic sequential data processing tasks. As compared to LSTM, RNN has a small scope of fine tuning variable components thus making it limited in certain aspects. Despite of these limitations RNN has it’s advantages when it comes to simple and effective architecture.

4.3 Support Vector Machine

SVM is a machine learning supervised algorithm. The focus of this algorithm is to draw a separating hyperplane between two classes. Based to this hyperplane the algorithm will determine in which class the input belongs to. While processing textual data the algorithm first converts the text to numbers in order to plot them on a graph. Those numbers can also be depicted as vectors being shown on graphs. Similar vectors are located close to each other, this makes the outlier detection easy. SVM is mostly effective for binary classification of data.

4.4 Logistic Regression

Logistic regression is a foundational machine learning algorithm. It is used for discrete classification unlike linear regression. Where linear regression predicts the probable output value, logistic regression outputs the class of prediction. The classification is done based on sigmoid function which determines the probability. Sigmoid curve can be plotted on a graph which differentiates between the two classes based on the threshold set in the model.

4.5 Random Forest

It is an ensemble learning algorithm and supervised in nature. The algorithm is based on constructing multiple decision trees during the training phase. And during the time of output the highest voted output amongst all decision trees is considered. The algorithm is preferred by many due to its less time consumption in the training phase and also a good accuracy score due to its ensemble nature.

Among the models covered, Recurrent Neural Networks show remarkable abilities to process sequential data and capture complex temporal dependencies. They effective tools in a variety of time-sensitive applications because of their capacity to efficiently store and use historical data, which enables the extraction of subtle patterns and trends within complicated datasets

The loss function used to evaluate the model is categorical crossentropy. The formula for categorical crossentropy is:

Categorical Crossentropy= $-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}.\log\left(p_{ij}\right)$

Here:

N is number of samples

C is the number of classes

y_{ij} is 1 if true class for sample i is class j, 0 otherwise.

p_{ij} is the predicted probability that sample i belongs to class j

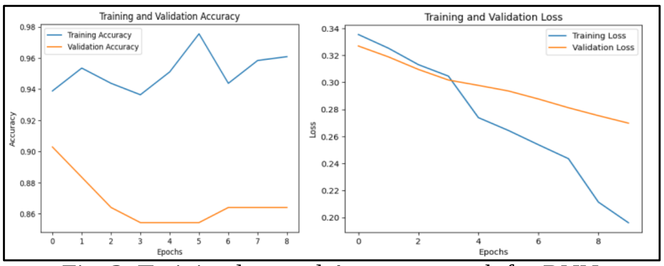


Fig. 3. Training loss and Accuracy graph for RNN

The loss and accuracy graph illustrates the progress made during the RNN deep learning model's training process. The model demonstrates effective learning ability, lowering errors during the training phase, with a comparatively low loss value of 0.419. The model has commendable accuracy of 0.81 which shows its capability of effective misinformation detection.

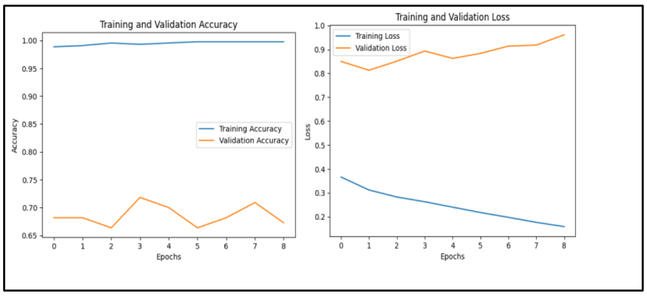


Fig. 4. Training loss and Accuracy graph for LSTM

The above graph shows training loss and accuracy graph for LSTM. The model has a very impressive learning curve in which it shows its effectiveness in processing textual data. The blue line represents training accuracy whereas orange represents validation accuracy.

RESULTS

Table. 2. Comparative Study

MODEL		CLASS	ACCURACY	PRECISION	RECALL	F1 - SCORE	SUPPORT	AUC-SCORE
ML	SVM	0	82	0.89	0.90	0.89	69	-
		1		0.90	0.88	0.89	68	
	LOGISTIC REGRESSION	0	80	0.85	0.88	0.87	69	
		1		0.88	0.84	0.86	68	
	RANDOM FOREST	0	79	0.86	0.80	0.83	69	
		1		0.81	0.87	0.84	68	
DL	LSTM	0	86	0.92	0.81	0.86	69	0.86
		1		0.83	0.93	0.88	68	
	RNN	0	82	0.91	0.72	0.80	88	0.82
		1		0.75	0.93	0.83	83	

Metrics including precision, recall, support, and f1 score were used to assess each implemented model. We were able to compare the implemented models thanks to the assessments' outcomes. After evaluating the outcomes, deep learning models fared better than machine learning models.

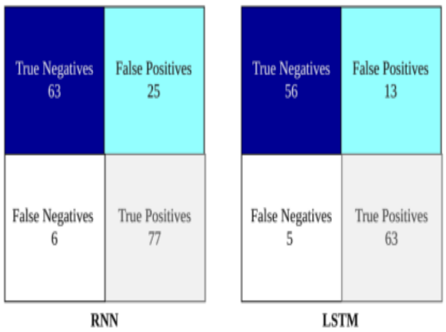


Fig. 5. Confusion Matrices

In order to calculate metrics like precision, recall and support we had to measure some raw parameters. These parameters give a qualitative measure of the outcomes the model gives. Those parameters indicate the amount of false classifications and true classifications of both the classes..

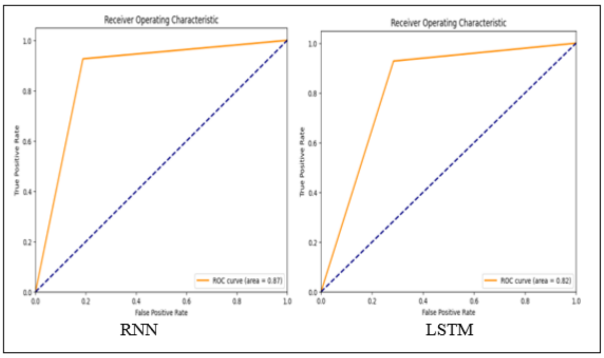


Fig. 6. AUC-ROC Curve

The above graph shows the AUC-ROC curve for LSTM and RNN models. The graph depicts the accuracy of models through the area present under the curve. The graph is plotted between correct and incorrect classifications of positive class.

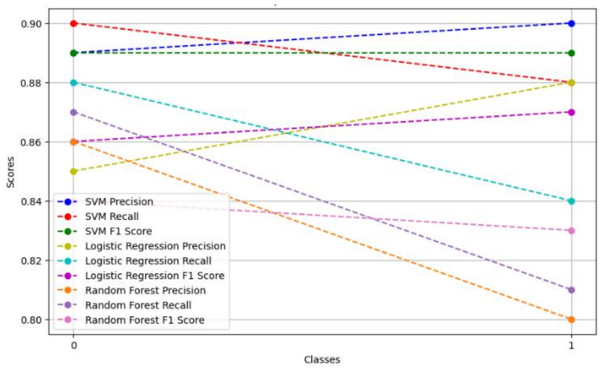


Fig. 7. Comparison of machine learning models

The above graph involves comparison between different machine learning models implemented in our study. X-axis represents the model names, whereas y-axis represents the evaluation scores. Every line in the graph represents a specific evaluation metric of a model. The different colors help differentiate between the various metrics and models.

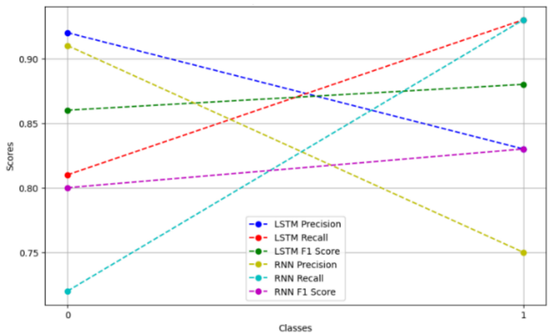


Fig. 8. Comparison of deep learning models

The above line graph compares the two deep learning models, namely LSTM and RNN. Every line in the graph represents some evaluation metric of the particular deep learning model. X-axis represents name of the model being compared whereas y-axis represents accuracy score of the model.

CONCLUSION

This research has explored diverse machine learning and deep learning algorithms with the purpose of unmasking misinformation in health domain The research was mainly focused on deep learning algorithm for achieving the objective of misinformation detection, the process of exploring these diverse algorithms demonstrated the reliability of RNN algorithm in detecting the misinformation. One of the major finding was that the RNN model was comparatively more efficient than other deep learning models which also comprised of LSTM. This also proves the ability of deep learning model to address crucial problems like misinformation in health domain. As for the future work the model can be enhanced by improving the quality of dataset upon which it is trained. The model needs to trained upon a very diverse dataset which is comprehensive enough to comprise the huge landscape of healthcare domain.

REFERENCES

[1] Enyan Dai, Yiwei Sun, Suhang Wang. 2020. Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository.

[2] Parfenenko, Yuliia & Verbytska, Anastasiia & Bychko, Dmytro & Shendryk, Vira "Application for Medical Misinformation Detection in Online Forums " 2020

[3] Sotto, Stefano & Viviani, Marco. "Health Misinformation Detection in the Social Web: An Overview and a Data Science Approach" 2022

[4] Jain, D.K., Kumar, A. & Shrivastava, A. "CanarDeep: a hybrid deep neural model with mixed fusion for rumour detection in social data streams" 2022

[5] A. Ghenai and Y. Mejova, "Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter" 2017

[6] G. Chaphekar and J. G. Jetcheva, "HealthLies: Dataset and Machine Learning Models for Detecting Fake Health News" 2022

[7] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, "Fake News Detection on Social Media: A Data Mining Perspective" 2017

[8] Zuhui Wang, Zhaozheng Yin, Young Anna Argyris "Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning" 2020

[9] Ning Xin Nyow and Hui Na Chua '"Detecting Fake News with Tweets Properties" 2020

[10] Yuliia Parfenenko ,Anastasiia Verbytska, Dmytro Bychko ,Vira Shendryk "Application for Medical Misinformation Detection in Online Forums" 2020

[11] Pritika Bahad , Preeti Saxena , Raj Kamal "Fake News Detection using Bi-directional LSTM Recurrent Neural Network" 2019

[12] Qing Liao, Heyan Chai, Hao Han, Xiang Zhang "An Integrated Multi-Task Model for Fake News Detection" 2017

[13] Maram Gharaibeh, Rasha Obeidat "Datasets and Approaches of COVID-19 Misinformation Detection" 2023

[14] M. F. Mridha, Ashfia Jannat, Keya, Md, Abdul Hamid "A Comprehensive Review on Fake News Detection With Deep Learning" 2017

[15] Min, Erxue & Ananiadou, Sophia. (2023)." PESTO: A Post-User Fusion Network for Rumour Detection on Social Media" 2023

[16] Meng, Weizhi & Kwok, Lam. "Adaptive False Alarm Filter Using Machine Learning in Intrusion Detection" 2022

[17] Zhang, Ying & Wang, Jianhui "Detecting False Data Injection Attacks in Smart Grids: A SemiSupervised Deep Learning Approach" 2021

[18] Hamdan, Yasir & Ammayappan, Sathesh. " Faultless Decision Making for False Information in Online" 2022

[19] Wang, Zuhui & Yin, Zhaozheng & Argyris, Young."Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning" 2020
[20] Saengkhunthod, Chotipong & Kerdnoonwong, Parischaya & Atchariya Chanwanich, Kanokwan. "Detection of Unreliable Medical Articles on Thai Websites" 2022
[21] Parfenenko, Yuliia & Verbytska, Anastasiia & Bychko, Dmytro & Shendryk, Vira "Application for Medical Misinformation Detection in Online Forums " 2020
[22] Ghasiya, Piyush & Okamura, Koji. " Investigating COVID-19 News across Four Nations: A Topic Modeling and Sentiment Analysis Approach" 2021
[23] Sotto, Stefano & Viviani, Marco. "Health Misinformation Detection in the Social Web: An Overview and a Data Science Approach" 2022
[24] Ivan Srba, Branislav Pecher, Matus Tomlein, Robert Moro, Elena Stefancova, Jakub Simko, Maria Bielikova. 2022. Monant Medical Misinformation Dataset: Mapping Articles to Fact-Checked Claims
[25] Neema Kotonya, Francesca Toni. 2020. Explainable Automated Fact-Checking for Public Health Claims