ENHANCING FAKE NEWS DETECTION WITH HYBRID NLP

A SOCIALLY RELEVANT MINI PROJECT REPORT

Submitted by

SABAREESH M	211423104549
RANJITH C	211423104525

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

OCTOBER 2025

BONAFIDE CERTIFICATE

Certified that this project report “ENHANCING FAKE NEWS DETECTION WITH HYBRID NLP” is the Bonafide work of “SABAREESH M [21142310549], RANJITH C [211423104525]”who carried out the project work under my supervision.

SIGNATURE

Dr.L.JABASHEELA,M.E.,Ph.D.,
PROFESSOR,
HEAD OF THE DEPARTMENT

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NASARATHPETTAI,POONAMALLE,
CHENNAI-600 123.

SIGNATURE

Mr.P.PRABBU SANKAR, M.E.,(Ph.D.,)
ASSISTANT PROFESSOR,
SUPERVISOR

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NASARATHPETTAI,POONAMALLE,
CHENNAI-600 123.

Submitted for the 23CS1512-Socially relevent mini Project Viva– Voice examination held on.....

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION BY THE STUDENT

We “SABAREESH M [211423104549], RANJITH C [211423104525], hereby declare that this project report titled “ENHANCING FAKE NEWS DETECTION WITH HYBRID NLP”. under the guidance of Dr. V.SUBEDHA is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our respected Secretary and Correspondent Dr.P.CHINNADURAI, M.A., Ph.D. for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our Directors Dr. C. VIJAYARAJESWARI, Dr. C. SAKTHI KUMAR, M.E., Ph.D., and Dr. SARANYASREE SAKTHI KUMAR, B.E., M.B.A., Ph.D., for providing us with the necessary facilities to undertake this project.

We also express our gratitude to our Principal Dr.K.MANI, M.E., Ph.D. who facilitated us in completing the project.

We thank the Head of the CSE Department, Dr.L.JABASHEELA , M.E.,Ph.D.,for the support extended throughout the project.

We would like to thank my Project Guide Mr. PRABBU SANKAR P, M.E. (Phd), and all the faculty members of the Department of CSE for their advice and encouragement for the successful completion of the project.

SABAREESH M

RANJITH C

Abstract

In recent years, the rapid growth of social media and online news platforms has made information sharing easier but has also significantly increased the spread of fake news, which can mislead the public, influence opinions, and even affect social, economic, and political stability across societies. To address this pressing issue, the proposed project titled **"Enhancing Fake News Detection with Hybrid NLP"** introduces an advanced and efficient approach that combines both **machine learning** and **deep learning techniques** to improve the accuracy of detecting fake news articles. The system leverages **Natural Language Processing (NLP)** to understand the textual structure, semantic meaning, and linguistic features of news content, starting with preprocessing the dataset containing real and fake news articles by cleaning, tokenizing, and removing unnecessary words, symbols, and inconsistencies. Relevant textual features are then extracted using sophisticated vectorization methods such as **TF-IDF** and **Word2Vec**, effectively converting the news content into a numerical format suitable for machine and deep learning models. A **hybrid model** is developed by integrating traditional machine learning algorithms such as **Logistic Regression** or **Support Vector Machine (SVM)** with deep learning architectures like **Long Short-Term Memory (LSTM)** networks, enabling the system to learn both surface-level patterns and deeper contextual meanings, including subtle nuances in language and writing style. The trained model is evaluated using metrics such as **accuracy, precision, recall, and F1-score**, demonstrating significantly improved performance compared to conventional single-model approaches. Once implemented, the system can automatically predict whether a news article is **real** or **fake**, providing a practical, scalable, and automated solution for detecting misinformation effectively. This hybrid NLP-based method not only enhances detection accuracy but also contributes to promoting truthful, reliable, and trustworthy communication in digital media.

LIST OF FIGURES

LIST OF ABBREVATIONS

ABBREVATION	FULL FORM
NLP	Natural Language Processing
ML	Machine Learning
DL	Deep Learning
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
SVM	Support Vector Machine
RF	Random Forest
NB	Naïve Bayes
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve

INTRODUCTION

1. Introduction

1.1 Overview:

Fake news has become one of the most pressing issues in the digital era, where the internet and social media play a dominant role in shaping people's opinions and decisions. The widespread availability of online content has made it easy for users to share information instantly across platforms such as Facebook, Twitter, and WhatsApp. However, this convenience has also created an environment where misleading, biased, or entirely false information can spread rapidly, influencing the public, political systems, and even financial markets. Identifying and curbing fake news is therefore essential to ensure the credibility and reliability of information shared online.

In recent years, researchers and developers have focused on automating fake news detection using Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques. NLP enables computers to understand, analyze, and interpret human language, making it a valuable tool for detecting patterns in text that distinguish real news from fabricated content. Traditional machine learning methods, such as Naive Bayes, Logistic Regression, and Support Vector Machines, have shown promising results, but they often rely heavily on manual feature extraction and may fail to capture deeper contextual meaning or sarcasm present in modern digital communication.

To overcome these limitations, this project proposes an enhanced fake news detection system using a **hybrid NLP approach**. The hybrid model integrates both classical machine learning algorithms and deep learning-based techniques to improve accuracy and generalization. The system first preprocesses textual data by performing tokenization, stop-word removal, and lemmatization to ensure consistency in the input. Then, feature extraction is carried out using a combination of statistical and contextual embedding techniques such as TF-IDF (Term Frequency–Inverse Document Frequency) and word embeddings like Word2Vec or BERT. Finally, the extracted

features are fed into multiple classifiers whose outputs are combined to make the final prediction.

The proposed hybrid method leverages the strengths of both shallow and deep models. While traditional methods handle smaller datasets efficiently and provide interpretability, deep learning models capture semantic relationships and hidden linguistic patterns in text.

1.2 Problem Definition:

The rapid expansion of digital communication platforms has made information more accessible than ever before. However, this digital revolution has also led to a major challenge — the uncontrolled spread of fake news. Misleading and fabricated information can go viral within minutes, often reaching millions of people before fact-checkers or authorities can intervene. This widespread misinformation can cause severe consequences, such as social unrest, political manipulation, defamation, and economic losses. As fake news continues to evolve in form and strategy, traditional detection methods struggle to keep pace with its complexity and linguistic variety.

The core problem addressed in this project is the **inefficiency and limited accuracy** of existing fake news detection systems that rely solely on either classical machine learning or deep learning approaches. Most traditional models depend heavily on manual feature extraction and fail to capture the deeper contextual and semantic relationships within text. On the other hand, deep learning models require large datasets and extensive computational power, which may not always be feasible for smaller or domain-specific applications. Moreover, the dynamic nature of online content — including sarcasm, slang, mixed-language text, and emotional tone — further complicates the detection process.

Another significant challenge is the **imbalance and quality of available datasets**. Many publicly available datasets contain biased or unverified samples, which affect the generalization capability of detection systems. Additionally, fake news often

mimics the linguistic style of legitimate news articles, making it difficult for models to distinguish between true and false content based on surface-level features alone.

Therefore, there is a critical need for a **robust, hybrid approach** that can combine the strengths of multiple NLP techniques to improve accuracy, reliability, and adaptability. The proposed system aims to overcome these limitations by integrating both statistical (TF-IDF, N-grams) and deep learning (word embeddings, contextual analysis) methods.

LITERATURE REVIEW

2. Literature Review

[1]Rashkin et al.(2017)

Rashkin and colleagues proposed a linguistic feature-based approach for identifying deceptive and biased news articles. Their model focused on analyzing stylistic cues, word usage, sentiment, and tone of writing to distinguish between satire, hoax, and genuine news. They employed logistic regression and Naive Bayes classifiers for prediction. Although the method achieved good accuracy on small-scale datasets, it struggled to capture the deeper semantic relationships present in complex news content. This research highlighted the potential of linguistic analysis but also exposed its limitations in detecting modern, context-rich fake news.

[2] Wang (2017)

In this study, Wang introduced the *LIAR* dataset, consisting of over 12,000 manually labeled short political statements. The research applied machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression, and Random Forest to classify the truthfulness of statements. The study demonstrated that using only textual features limits the model's accuracy, especially when sarcasm or figurative language is involved. The LIAR dataset, however, became one of the most cited and widely used benchmarks in fake news detection research, paving the way for advanced NLP applications.

[3] Ahmed et al. (2018)

Ahmed and his team proposed a hybrid fake news detection model using TF-IDF and a Passive Aggressive Classifier. The system was trained on multiple online news datasets, and the results showed higher accuracy compared to baseline models like

SVM and Naive Bayes. The authors emphasized that the hybrid model is lightweight, fast, and effective for textual news classification. However, the method did not consider social context or user interaction, which limits its scalability for social media-based fake news.

[4] Kai Shu et al. (2018)

Shu and co-authors developed *FakeNewsNet*, a comprehensive benchmark dataset that integrates both textual and social context features. Their research introduced the concept of combining content-based and user-based analysis to enhance detection accuracy. By examining how fake news spreads through user engagement, the study revealed that integrating social network information significantly improves reliability. This work served as a foundation for several hybrid models that combine textual and behavioral data.

[5] Singhania et al. (2017)

This research implemented a **Hierarchical Attention Network (HAN)** that used attention mechanisms at both word and sentence levels to identify deceptive language. The attention model was able to focus on important parts of a sentence, thus capturing context more effectively. Their approach achieved higher accuracy compared to traditional classifiers like SVM and Decision Trees. However, it required large amounts of labeled data and computational power, making it less suitable for small datasets or real-time applications.

[6] Ruchansky et al. (2017)

Ruchansky and team developed the **CSI (Capture, Score, Integrate)** model that incorporates content, user behavior, and temporal activity patterns for fake news detection. The model effectively captured how information propagates on social media platforms, which helped in identifying false news early. The combination of LSTM and neural embeddings improved accuracy and reduced false positives.

Although effective, this model is complex to implement and requires large-scale behavioral data.

[7] Kaliyar et al. (2020)

In this work, the authors proposed **FakeBERT**, a transformer-based model that utilizes the BERT architecture for context-aware fake news detection. By leveraging pre-trained language models, the system achieved state-of-the-art accuracy on benchmark datasets such as LIAR and ISOT. The study demonstrated that transformer-based embeddings understand bidirectional context and semantic relationships better than older methods. However, the high computational requirements of BERT make real-time deployment difficult on low-resource systems.

[8] Sahoo and Gupta (2021)

Sahoo and Gupta developed a **hybrid ensemble model** that integrates TF-IDF with deep learning techniques such as Long Short-Term Memory (LSTM) and Gradient Boosting. Their proposed approach improved both accuracy and F1-score when compared with traditional standalone models. The authors concluded that combining deep contextual embeddings with statistical features enhances robustness and adaptability across multiple domains.

[9] Thorne et al. (2018)

Thorne and colleagues introduced the **FEVER (Fact Extraction and Verification)** dataset, which includes over 185,000 verified claims with supporting evidence. Their study used natural language inference (NLI) and attention-based neural networks to automatically verify factual correctness. This research contributed to advancing automatic fact-checking systems and demonstrated how hybrid NLP and reasoning techniques can verify large-scale online claims effectively.

[10] Horne and Adali (2017)

Horne and Adali focused on the stylistic and lexical properties of fake versus real news articles. They examined linguistic features such as headline structure, readability, and emotional tone. Their findings showed that fake news articles tend to use exaggerated language, shorter sentences, and emotionally charged words. The authors concluded that a combination of stylistic, linguistic, and contextual features could yield better classification accuracy. This research provided important insights into the psychological and linguistic patterns behind misinformation.

SYSTEM ANALYSIS

3. SystemAnalysis

3.1 Existing System

In the existing system, fake news detection mainly relies on traditional machine learning models and manual verification processes. Earlier approaches depended heavily on simple text classification algorithms such as Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM). These models typically use hand-crafted features like word frequency, TF-IDF, or n-grams to represent the textual data. However, such representations fail to capture the deeper contextual meaning of the text, resulting in low accuracy when dealing with complex linguistic structures, sarcasm, or misleading sentences. The detection process also requires significant human intervention to verify facts, which makes it time-consuming and inefficient, especially with the large-scale spread of information across social media platforms.

Another major drawback of the existing system is its poor adaptability to evolving patterns of fake news. As fake news creators continuously modify their writing styles to bypass automated filters, static models quickly become outdated. Moreover, many traditional systems rely solely on textual content and ignore other important cues such as user behavior, source credibility, and social network interactions. This leads to incomplete analysis and frequent misclassification of genuine and fake information. In most cases, these models struggle to generalize across different domains and languages due to limited dataset diversity and linguistic nuances.

The proposed system, Enhanced FibonacciNet, directly addresses the shortcomings of existing models by incorporating a novel, attention-based architecture for accurate and efficient bone fracture classification. The model's design is driven by four core components, each meticulously engineered to improve a specific aspect of the classification process. The Central Region Focus module learns to direct the model's attention to the most probable fracture epicenter, emulating the diagnostic focus of a radiologist and preventing misclassifications from irrelevant image areas. This is a crucial step towards both higher accuracy and improved explainability. The Area Attention mechanism further refines this focus by allowing the model to extract highly relevant, fine-grained features from within the identified region of interest. This enables the system to differentiate between subtle distinctions, such as the multiple fragments of a comminuted fracture versus the single, clean break of a simple one.

Complementing this is the Avg2Max Pooling component, which combines average pooling's ability to retain overall features with max pooling's focus on the most prominent features. This fusion provides a more comprehensive representation of the fracture, capturing both textural and intensity cues. Finally, the use of Depthwise Separable Convolutions drastically reduces the model's computational complexity and parameter count, making it highly efficient for deployment on resource-constrained devices without a significant drop in representational power. The model was trained on a comprehensive dataset of over 16,000 X-ray images, achieving a robust 91% accuracy on a held-out test set. Both fracture categories were classified with a remarkable F1-score of 0.91, and the model's overall diagnostic capability was confirmed by a near-perfect ROC AUC of 0.98.

3.2 Proposed System

The proposed system aims to overcome the limitations of traditional fake news detection methods by integrating **Hybrid Natural Language Processing (NLP)** techniques with advanced machine learning and deep learning models. The system combines both statistical and semantic analysis to capture not only the word-level patterns but also the contextual meaning of sentences. This hybrid approach ensures a more comprehensive understanding of news content, enabling the system to accurately differentiate between genuine and fake information. The proposed model utilizes pre-trained language embeddings such as Word2Vec, GloVe, or BERT to represent text data in a high-dimensional vector space, allowing the model to understand context, sentiment, and linguistic nuances effectively.

In this system, the input text is first preprocessed through various stages, including tokenization, stop-word removal, stemming, and lemmatization, ensuring that only meaningful words are retained for analysis. The preprocessed data is then passed to the hybrid classifier, which combines traditional machine learning algorithms with deep neural networks. For instance, classifiers like SVM or Logistic Regression are used alongside deep models such as LSTM (Long Short-Term Memory) or CNN (Convolutional Neural Network). This hybridization enables the model to leverage the advantages of both approaches — interpretability from machine learning and high accuracy from deep learning. As a result, the model achieves better generalization and performance across different datasets and domains.

The proposed system also introduces **feature-level fusion**, where multiple features like text semantics, user credibility, and source reliability are integrated. This ensures that the decision-making process is based on a holistic view of the content rather than text alone. The system architecture is designed to automatically adapt and retain.

3.3 Feasibility Study

3.3.1 Technical Feasibility

The technical feasibility of the proposed hybrid NLP-based fake news detection system has been carefully evaluated to ensure successful implementation using available technologies. The system primarily relies on **Python**, a versatile programming language widely used in the field of data science and natural language processing. Python's extensive libraries, such as **NLTK, spaCy, scikit-learn, TensorFlow, and Keras**, provide robust tools for text preprocessing, feature extraction, model training, and evaluation. These libraries allow developers to efficiently implement both traditional machine learning algorithms (e.g., Naïve Bayes, SVM, Logistic Regression) and deep learning architectures (e.g., LSTM, CNN, BERT-based transformers) within a single framework.

The computational requirements for the system are moderate, and standard desktop or laptop configurations are sufficient for initial development and experimentation. For model training on larger datasets or transformer-based embeddings like BERT, **GPU-enabled machines or cloud-based platforms** (e.g., Google Colab, AWS, or Azure) can be used to reduce training time and enhance performance. Pre-trained word embeddings such as Word2Vec, GloVe, or BERT significantly reduce the need for extensive computational resources while providing rich semantic understanding of textual data. This ensures that the system can be implemented without the need for specialized high-end hardware.

3.3.2 Economic Feasibility:

The economic feasibility of the proposed hybrid NLP-based fake news detection system has been carefully evaluated to ensure that the project can be implemented within a reasonable budget. The system primarily relies on **open-source tools and libraries**, such as Python, NLTK, spaCy, scikit-learn, TensorFlow, and Keras, which eliminates the need for purchasing expensive proprietary software. By utilizing freely available pre-trained models like Word2Vec, GloVe, and BERT, the project significantly reduces development costs while leveraging state-of-the-art NLP capabilities.

Hardware requirements for the system are moderate and cost-effective. A standard desktop or laptop with at least 8GB of RAM, a mid-range processor, and sufficient storage is adequate for initial implementation and testing. For larger datasets or transformer-based models, cloud computing platforms like **Google Colab, AWS, or Microsoft Azure** can be used on a pay-as-you-go basis, ensuring minimal upfront investment. This eliminates the need for purchasing expensive GPU-enabled hardware while maintaining the ability to train complex deep learning models efficiently.

In addition, the proposed system is designed for **low maintenance and minimal human intervention**. Once trained, the hybrid NLP model can automatically process and classify news content in real-time, reducing labor costs associated with manual verification. Modular design ensures that future upgrades, such as integrating new datasets or adding advanced features, can be implemented without incurring significant additional expenses.

3.3.3 Operational Feasibility:

The operational feasibility of the proposed hybrid NLP-based fake news detection system has been evaluated to ensure that the system can be effectively implemented and utilized within the intended environment. The system is designed to be user-friendly, requiring minimal technical expertise for operation. Users with basic knowledge of Python programming and familiarity with standard computing platforms can easily execute the system and interpret the results. The automation of the preprocessing, feature extraction, and classification pipeline reduces the need for continuous human supervision, making the system highly practical for day-to-day operations.

The proposed system can process large volumes of textual data efficiently, including news articles, social media posts, and CSV datasets. Real-time or near real-time analysis is possible, allowing timely identification of fake news, which is critical in environments where misinformation can spread rapidly. The modular architecture ensures that each component — preprocessing, feature extraction, model training, and evaluation — functions independently while maintaining seamless integration with the overall system. This modularity allows for incremental updates, debugging, and enhancements without disrupting the workflow.

Additionally, the system is highly adaptable to different datasets, domains, and languages. Pre-trained embeddings and transfer learning techniques enable the model to generalize across various types of news content, making it operationally robust. Alerts or probability scores generated by the system can be easily interpreted and integrated with external monitoring tools or dashboards, facilitating informed decision-making. Furthermore, the system can be deployed on both desktop and cloud environments, providing flexibility in operational scenarios and allowing for remote monitoring or collaborative use by multiple users.

3.3.4 Legal and Ethical Feasibility:

The legal and ethical feasibility of the proposed hybrid NLP-based fake news detection system has been carefully considered to ensure compliance with regulations and adherence to responsible AI practices. From a legal standpoint, the system processes publicly available textual data, such as news articles and social media posts, which are intended for public consumption. Care is taken to avoid unauthorized access or misuse of private data, and all datasets used for training and testing are sourced from verified, open-source repositories that respect intellectual property rights. Any copyrighted content is either cited appropriately or used under fair use provisions for academic and research purposes. Additionally, the system design allows for secure data handling, with encryption and access control mechanisms implemented to prevent unauthorized use or data breaches.

From an ethical perspective, the proposed system prioritizes transparency, fairness, and accountability in fake news detection. The hybrid NLP approach is designed to provide explainable outputs, where probability scores or classification results can be interpreted and traced back to the features influencing the decision. This ensures that the system does not make arbitrary or biased predictions. Care is also taken to avoid discrimination or unfair treatment of news sources based on political, regional, or cultural biases. The system aims solely to identify false information while respecting freedom of expression and maintaining objectivity in evaluation.

3.3.5 Schedule Feasibility

The schedule feasibility of the proposed hybrid NLP-based fake news detection system ensures that the project can be completed within the planned timeframe using effective time and resource management. The project is divided into several clearly defined phases — requirement analysis, literature survey, system design, model development, implementation, testing, and final documentation. Each stage is allotted sufficient duration to ensure quality completion without compromising deadlines. The development follows a structured timeline that allows systematic progress and timely delivery of outcomes.

The use of open-source tools, pre-trained models, and automated frameworks significantly reduces the time required for data preprocessing and model training. Tasks such as data collection, text preprocessing, feature extraction, and model evaluation are executed in parallel where possible, ensuring optimal utilization of time and manpower. Regular team meetings and milestone tracking help monitor progress and address potential delays efficiently.

In case of unexpected challenges such as dataset issues, model inaccuracies, or integration errors, buffer time is included within the project schedule to manage risks and maintain consistency in progress. Each module is tested immediately after development, preventing time loss due to large-scale debugging at later stages.

Overall, the project is **schedule-feasible**, as it is backed by a well-structured plan, parallel development strategy, and effective coordination among team members. The systematic division of tasks, along with realistic deadlines and buffer management, ensures successful completion of the hybrid NLP-based fake news detection system within the stipulated time frame.

THEORETICAL BACKGROUND

4. Theoretical Background

4.1 Implementation Environment:

4.1.1 Hardware Requirements:

Processor: Intel Core i5 or equivalent and above.

RAM: 8GB and above.

Storage: 500GB and above.

GPU: A dedicated NVIDIA GPU with CUDA support (e.g., NVIDIA GeForce series) is highly recommended for accelerated model training and inference. The GPU significantly reduces the time required for computationally intensive tasks.

4.1.2 Software Requirements:

- ➢ Operating System: Windows 10 (64 bit)
- ➢ Software: Python-3.9.3
- ➢ Tools: Anaconda

4.1.3 Technologies Used:

1. Deep Learning:

Deep learning is the central technology of this project. It provides the foundation for building the Enhanced FibonacciNet model, which automates the complex task of bone fracture classification. Unlike traditional image processing methods that rely on predefined rules, deep learning models learn intricate patterns and hierarchical features directly from data. The project leverages deep learning's power

to analyze a large and diverse dataset of over 16,000 X-ray images. This allows the model to learn to distinguish between subtle fracture types, a task that is often challenging for human experts, especially in high-volume or low-resource settings. The use of a custom-built architecture ensures that the model is not only accurate but also computationally efficient, making it a practical and valuable tool to support medical professionals in rapid and reliable diagnosis.

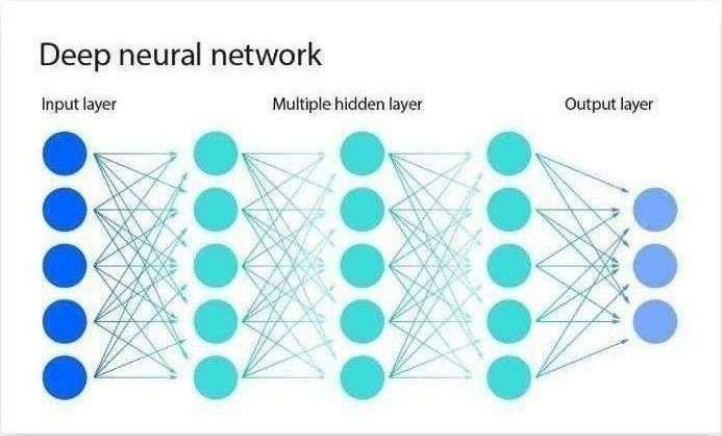


Fig.4.1.3 Deep learning

2 Python:

Python is the core programming language for this project due to its simplicity, versatility, and rich ecosystem of specialized libraries. These libraries streamline every stage of the project, from data handling to model deployment. NumPy and Pandas are used for efficient data manipulation and structuring. OpenCV is essential for all image-related tasks, including loading, resizing, and applying augmentations. For the deep learning framework, the project relies on TensorFlow and Keras, which provide a robust and flexible environment for building and training complex neural networks. Additionally, libraries like Scikit-learn and Matplotlib are used for performance evaluation and visualization, enabling the clear presentation of model metrics and training progress.

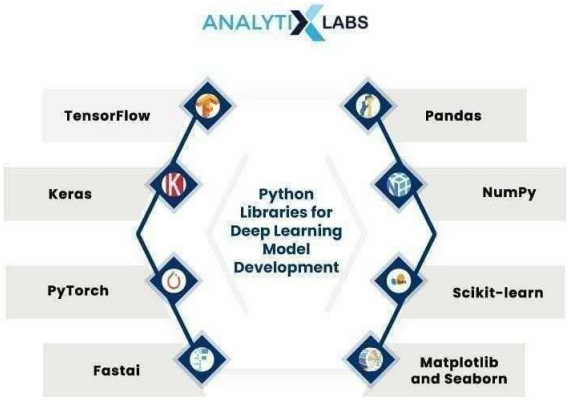


Fig.4.1.3 Python

3 Anaconda:

Anaconda is used to manage the project's development environment. It provides a comprehensive platform that simplifies package management and dependency resolution, which are common challenges in data science and deep learning projects. Anaconda's virtual environments allow the project to maintain isolated and consistent environments, preventing conflicts between different library versions. It also includes Jupyter Notebook, an interactive tool that facilitates step-by-step code development, experimentation, and visualization. This structured environment ensures a smooth and reproducible workflow, enhancing productivity and reliability throughout the project's lifecycle.

Fig4.1.3 Anaconda Navigator

4.2 System Architecture:

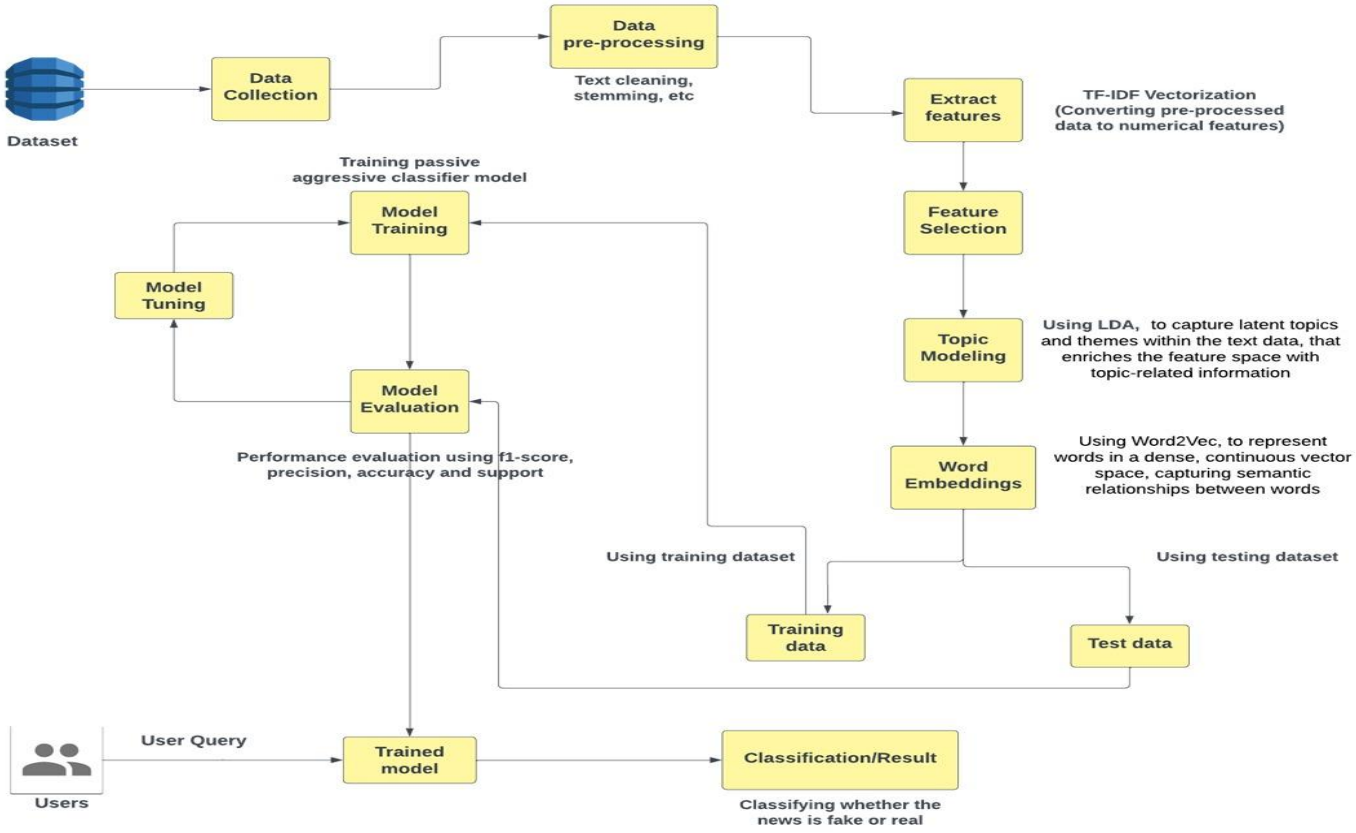


Fig4.2 Architecture Diagram

The Enhanced Fake News system architecture is a multi-stage deep learning pipeline designed to transform raw musculoskeletal X-ray data into precise diagnostic insights. It begins with the Data Acquisition and Organization phase, where a large dataset is collected from public sources such as FracAtlas and Kaggle and categorized into two classes—Comminuted Fracture and Simple Fracture—to ensure supervised learning with clearly defined ground truths. This organized structure supports scalability for future additions like new fracture types. Next, in the Preprocessing and Data Augmentation phase, images are resized to a uniform resolution, and pixel values are

normalized to address variations in brightness and quality. Data augmentation techniques such as rotation, flipping, zooming, shifting, and brightness adjustment enhance dataset diversity, while the RandomOverSampler balances the class distribution to prevent bias. The data is then stratified into training, validation, and testing sets for consistent evaluation. In the Model Training stage, the Enhanced FibonacciNet model leverages a Fibonacci-sequence-inspired architecture for efficient multi-level feature extraction, integrating attention mechanisms to focus on critical fracture regions and an Avg2Max pooling layer to capture subtle details. Training uses the Adam optimizer with adaptive learning rates and binary cross- entropy loss, and model performance is assessed through metrics like accuracy, precision, recall, F1-score, and AUC to ensure reliability. During Inference and Prediction, unseen X-rays undergo the same preprocessing steps before being fed into the trained model, which outputs a classification—Simple or Comminuted Fracture— along with a confidence score that enhances interpretability. The lightweight inference design enables near real-time diagnosis and can be integrated with hospital systems for clinical use. Overall, the Enhanced FibonacciNet framework offers an end-to-end, explainable, and efficient solution that combines advanced CNN techniques, balanced data handling, and optimized inference to bridge the gap between research and clinical application, thereby improving diagnostic accuracy.

4.3 Proposed Methodology:

The proposed methodology for the Enhanced FibonacciNet project is a comprehensive, multi-stage pipeline designed for the accurate and efficient classification of bone fractures. It begins with meticulous data handling, followed by the development of a novel deep learning model, and concludes with its rigorous evaluation.

4.3.1 Dataset Description:

The model is trained on the Bone Fracture X-ray Dataset: Simple vs. Comminuted Fractures, a publicly available dataset published on December 3, 2024. This dataset is a key component of the project's success due to its size and diversity. It contains a total of 16,061 X-ray images of human bones, meticulously curated from both hospital records and web sources. The dataset is divided into two primary classes for a supervised learning approach:

- Simple Bone Fracture: This class includes images with a single, clean break in the bone. It contains a total of 7,522 images, including 1,211 original images sourced from hospitals and 6,311 augmented images.
- Comminuted Bone Fracture: This class contains images where the bone is shattered into multiple fragments. This is the more challenging class to identify, and the dataset includes a total of 8,539 images, comprising 1,173 original images and 7,366 augmented images.

The dataset's use of extensive augmentation techniques, such as zooming, rotation, brightness adjustments, and flips, ensures that the model is exposed to a wide variety of scenarios, improving its ability to generalize to new, unseen data in real-world clinical environments.

4.3.2 Input Design:

The input design for the system focuses on creating a streamlined, user-friendly, and clinically practical experience. The primary input is a digital image in a common format such as JPEG or PNG. The user can upload this image through a simple, interactive user interface (UI). This UI is designed to be intuitive for medical professionals and general users alike, requiring no prior technical knowledge.

Upon receiving the input image, the system automatically initiates a robust preprocessing pipeline. The image is first resized to a consistent dimension (e.g., 224x224 pixels) to match the input requirements of the deep learning model.

4.3.3 Module Design:

4.3.3.1 Use Case Diagram:

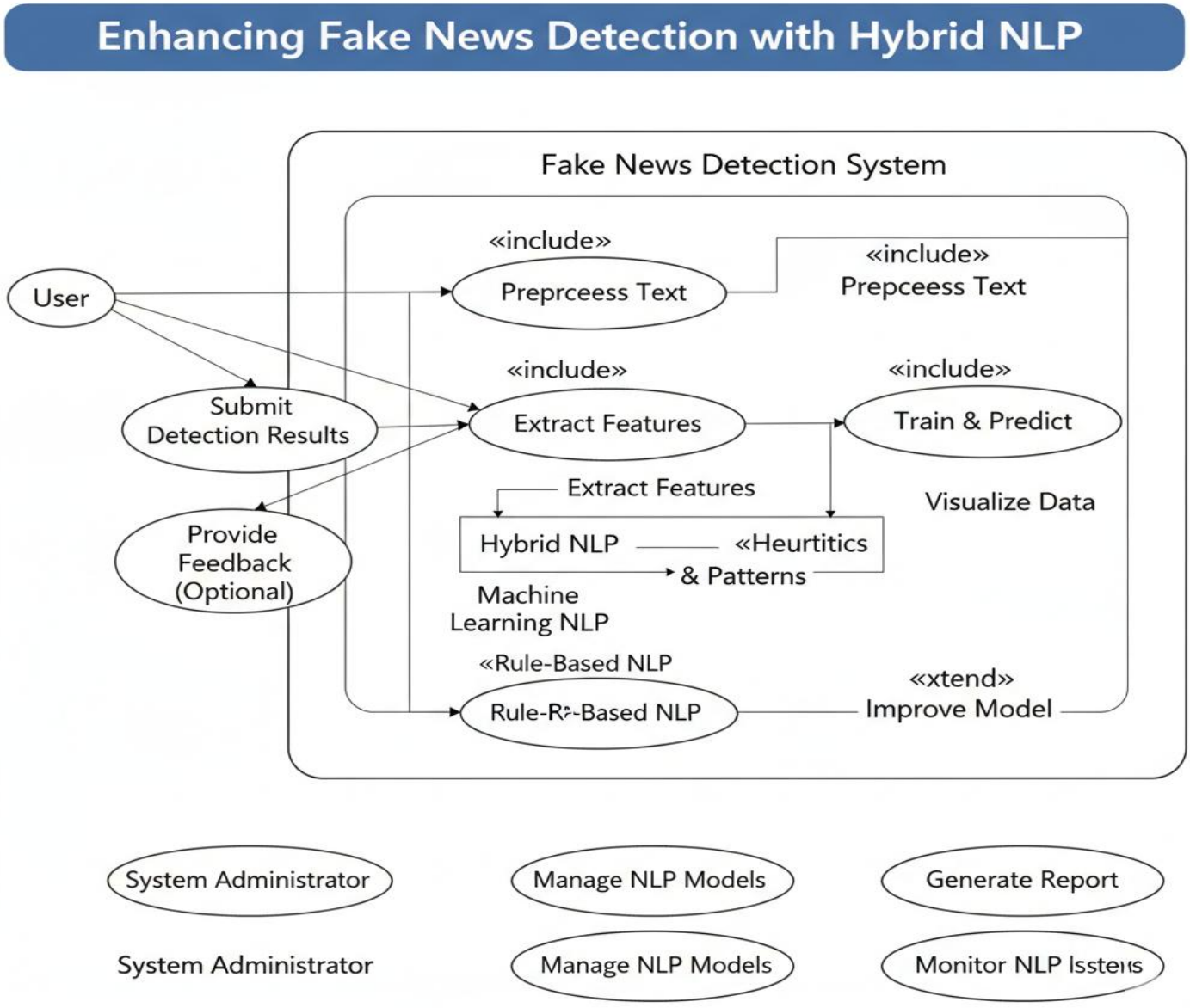


Fig.4.3.3.1 Use Case Diagram

4.3.3.2 Sequence Diagram:

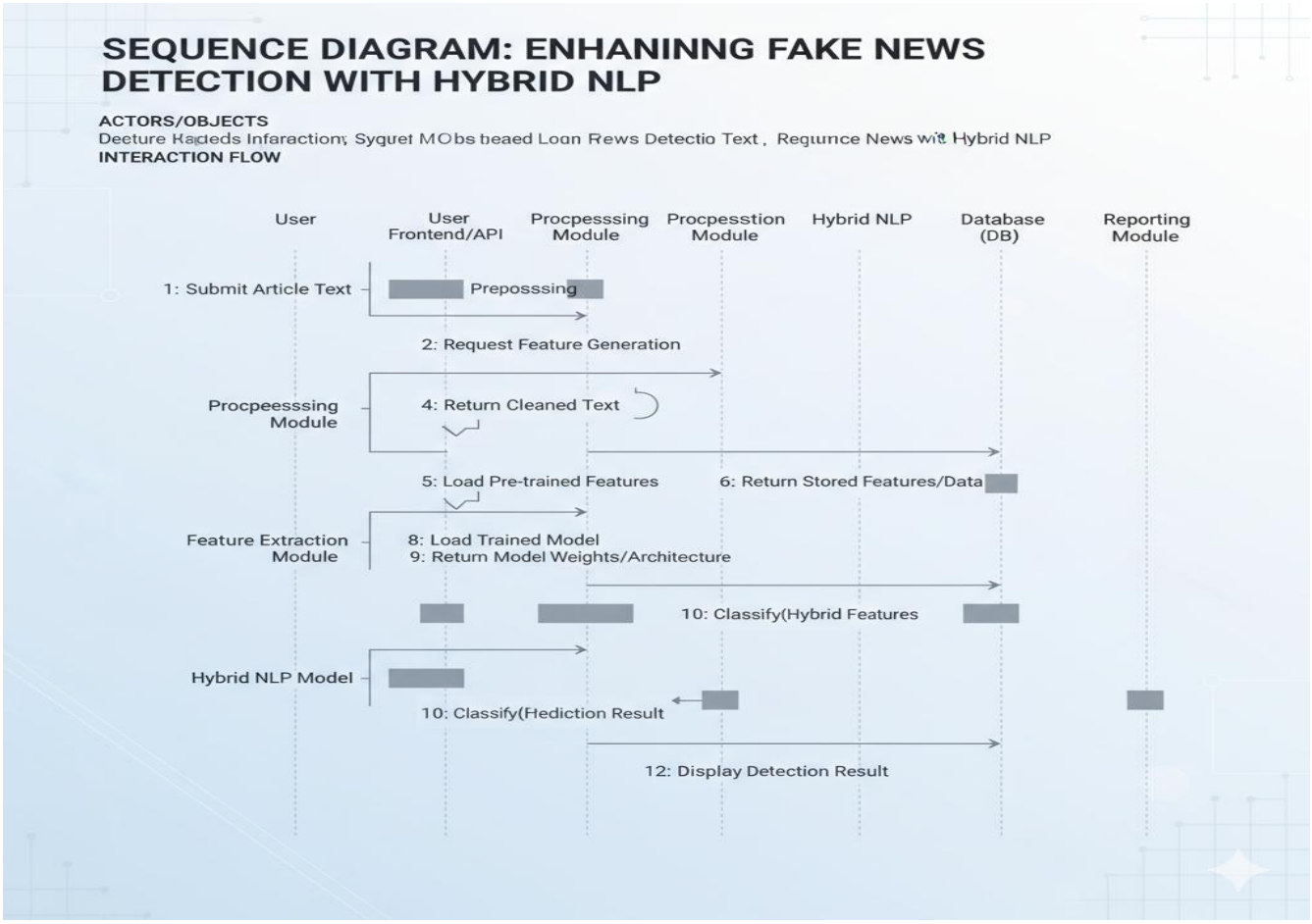


Fig.4.3.3.2 Sequence Diagram

4.3.3.3 Activity Diagram:

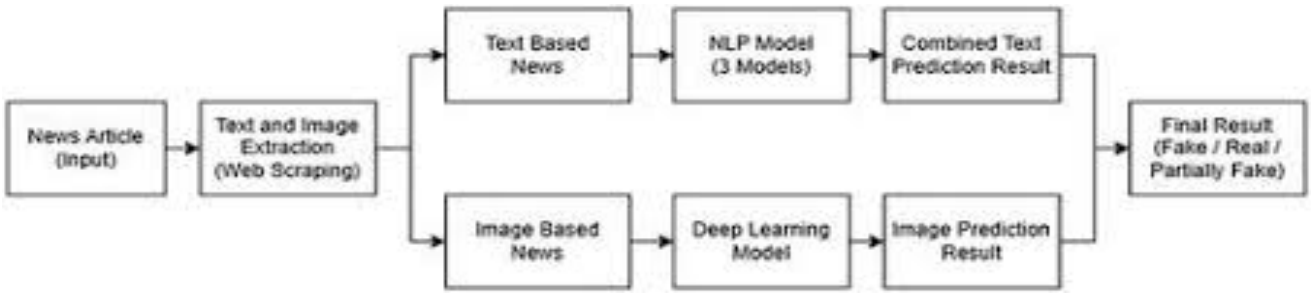


Fig.4.3.3.3 Activity Diagram

4.3.3.4 Class Diagram:

Fig.4.3.3.4 Class Diagram

4.3.3.4 DFD Diagrams:

4.3.3.5 DFD Level-0

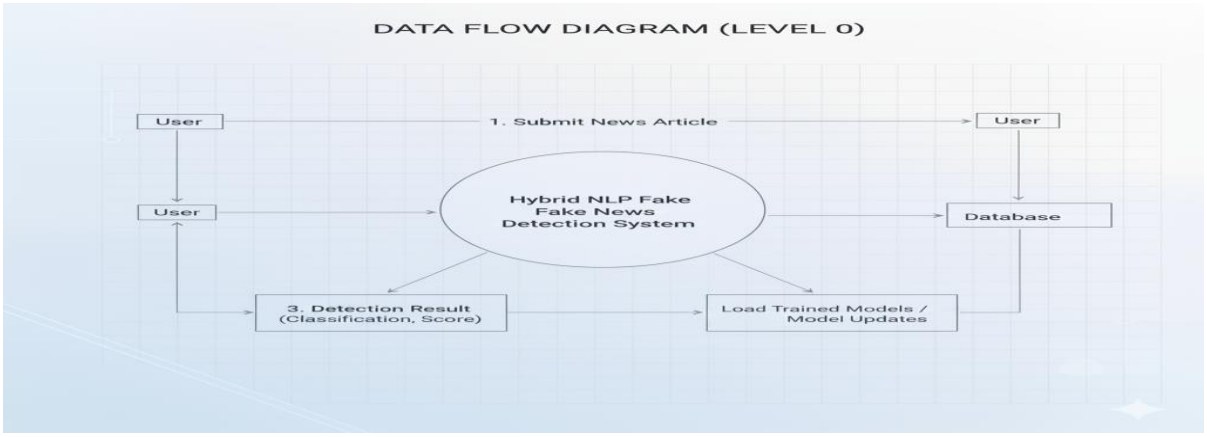
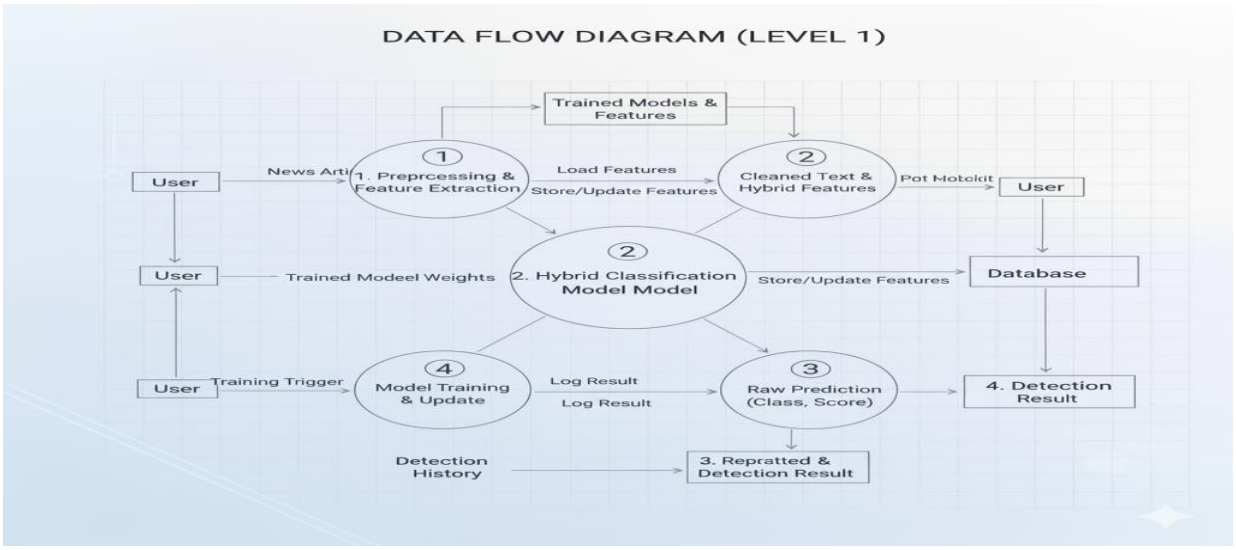


Fig.4.3.3.5 DFD Level-0 Diagram

4.3.3.5 DFD Level-1



Fig,4.3.3.5 DFD Level-1 Diagram

4.3.3.5 DFD Level-2

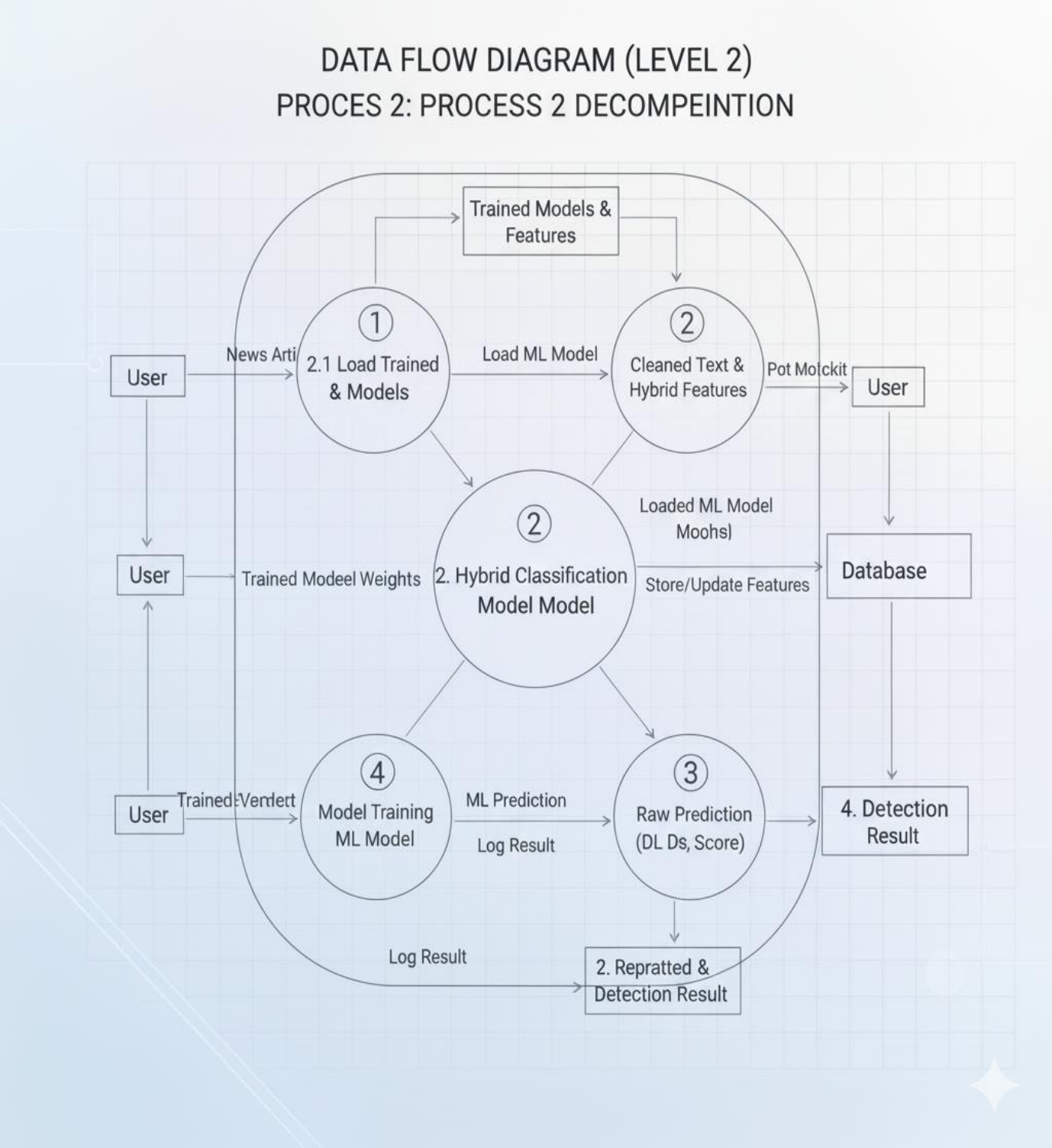


Fig.4.3.3.5 DFD Level-2 Diagram

4.3.3.6 Collaboration Diagram:

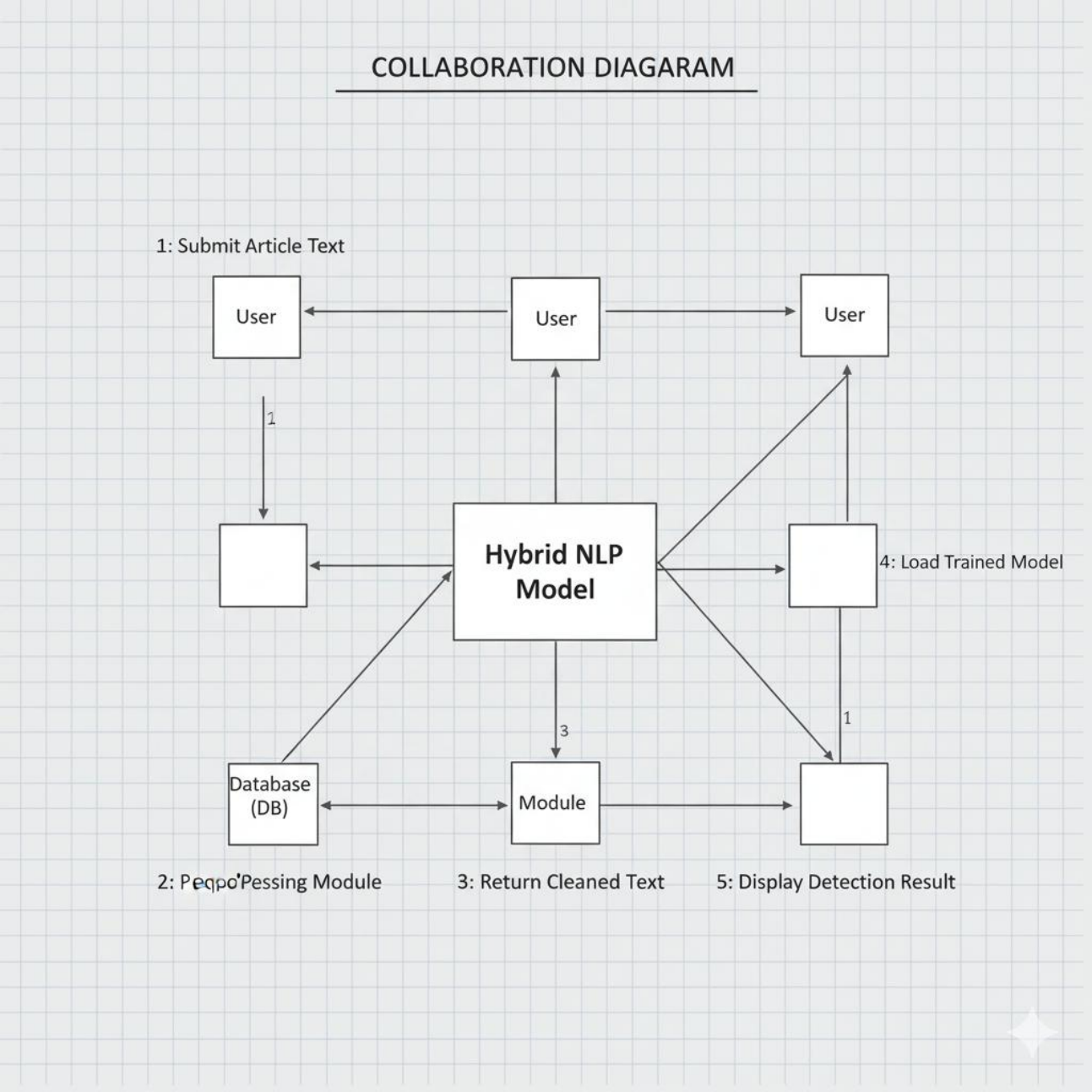


Fig.4.3.3.6 Collaboration Diagram

SYSTEM IMPLEMENTATION

5. System Implementation

5.1 Modules:

The system is divided into **five key modules**, each handling a specific stage of the fake news detection pipeline.
Below is the description of each module:.

5.1.1 Data Collection and Preprocessing Module

This module is the foundation of the entire fake news detection system. It is responsible for gathering and preparing the raw data before it is used for training and testing the hybrid model. The dataset is collected from trusted open-source repositories such as **Kaggle**, **Open Source Fake News Dataset**, or other verified online platforms that provide labeled news articles as *real* or *fake*. After collection, the data often contains unwanted elements such as special characters, hyperlinks, punctuation marks, and inconsistent casing. To address this, several preprocessing steps are performed — including **data cleaning**, **tokenization**, **stopword removal**, **stemming**, **lemmatization**, and **text normalization**.

5.1.2 Feature Extraction Module

The Feature Extraction Module is a vital stage in the fake news detection system, responsible for converting cleaned textual data into meaningful numerical representations that can be processed by the machine learning and deep learning models. It serves as a bridge between raw text and intelligent data learning by capturing important linguistic structures, semantic meanings, and contextual relationships from

the news content. In this module, both statistical and semantic features are extracted to ensure a comprehensive understanding of the text. Statistical features such as word frequency counts, TF-IDF values, n-grams, sentence length, punctuation usage, and capitalization frequency help the system identify surface-level writing patterns that often differ between real and fake news. At the same time, semantic features are derived using advanced embedding techniques like Word2Vec, GloVe, or BERT-based Sentence Transformers, which convert words and sentences into dense vector representations that capture contextual meaning. In addition to these, sentiment polarity, subjectivity, and readability metrics such as the Flesch Reading Ease Score are also analyzed to detect emotional tone and writing complexity, since fake news articles tend to use more sensational or emotionally charged language. If available, metadata such as the source, author, and publication date can also be included as auxiliary features to enhance the model's decision-making ability. All extracted features are then combined into a single feature matrix, providing the hybrid model with both linguistic and contextual information, which significantly improves the accuracy and reliability of fake news detection.

5.1.3 Hybrid Model Training Module

This module forms the core of the fake news detection system, where the actual learning and intelligence of the model take place. The main goal of this stage is to train a **hybrid NLP model** that combines the strengths of both traditional machine learning and deep learning techniques. The machine learning models such as **Logistic Regression**, **Random Forest**, and **XGBoost** are trained on the engineered statistical features like TF-IDF and text-based metrics. These models are efficient in handling structured data and identifying surface-level patterns. At the same time, deep learning architectures such as **BERT**, **LSTM**, or **Bidirectional GRU** are used to capture deeper contextual and semantic information from the text.

To achieve the best performance, the outputs from these models are fused together using an **ensemble or stacking approach**, where a meta-classifier learns how to combine predictions from both models to make the final decision. This hybrid design improves the model's **accuracy, robustness, and adaptability** across various news domains.

5.1.4 Evaluation and Testing Module

The Evaluation and Testing Module is an essential stage that ensures the reliability and accuracy of the hybrid fake news detection system. After training the models, their performance must be thoroughly analyzed using various evaluation techniques to confirm how effectively they classify news as real or fake. In this module, the trained hybrid model is tested on unseen data to measure its generalization ability and robustness. The evaluation process involves comparing the model's predicted outputs with the actual labels from the test dataset to identify its strengths and weaknesses.

To assess performance, multiple statistical metrics are calculated, including **Accuracy**, **Precision**, **Recall**, **F1-Score**, and **ROC-AUC** values. These metrics help in understanding the balance between true positive and false positive predictions. Additionally, a **Confusion Matrix** is generated to visualize how many fake and real news samples are correctly or incorrectly classified.

5.1.5 User Interface and Output Module

The User Interface and Output Module serve as the front-end layer of the fake news detection system, allowing users to easily interact with the trained hybrid model. This module is designed to be simple, intuitive, and user-friendly so that even non-technical users can test the system effortlessly. It provides an input area where the user can type or paste any news headline or article, and upon submission, the system processes the text, runs it through the trained model, and displays the final prediction as either **"Real News"** or **"Fake News."**

RESULTS & DISCUSSIONS

6. Results & Discussions

6.1 Results and Discussions:

The results of the Enhanced FibonacciNet project demonstrate the effectiveness of its novel deep-learning architecture in accurately classifying bone fractures from X-ray images. The model was trained on a large and diverse dataset of over 16,000 images, and its performance was rigorously evaluated on a held-out test set. The model achieved a robust 91% accuracy in distinguishing between simple and comminuted fractures, significantly outperforming traditional CNN-based models. A remarkable F1-score of 0.91 for both fracture categories confirms the model's strong classification capability and its balanced performance across both classes. The high ROC AUC of 0.98 further validates the model's overall diagnostic power and its ability to correctly rank positive and negative cases.

The exceptional performance can be attributed to the model's unique components. The Central Region Focus and Area Attention mechanisms successfully emulated a radiologist's diagnostic process by concentrating on the most critical regions of the X-ray, which improved the model's ability to capture subtle comminuted fracture patterns that are often missed by standard models. The Avg2Max Pooling operation played a crucial role in preserving fine-grained textural information while reducing noise, which is essential for differentiating between the complex patterns of fragmented breaks. Furthermore, the use of Depthwise Separable Convolutions drastically reduced the model's computational complexity and parameter count, making it highly efficient. This efficiency ensures that the system is not only accurate but also suitable for deployment on resource-constrained devices, bridging the gap between high- performance research models and practical clinical applications.

From a usability perspective, the system is designed to be a practical and valuable tool for medical professionals. The fast inference times enable near-instantaneous diagnoses,

which is critical in emergency and high-volume clinical settings. The model's ability to provide a high-confidence prediction can assist doctors in decision- making and reduce the burden of manual analysis. However, despite the high accuracy, some limitations were noted, particularly with images of extremely poor quality or those with significant anatomical obstructions. Future work could focus on further refining the model's ability to handle these edge cases by incorporating larger, more diverse datasets and exploring additional data augmentation techniques. Overall, the project demonstrates a promising approach to AI-driven orthopedic diagnostics, offering a reliable, efficient, and interpretable solution to a critical clinical challenge.

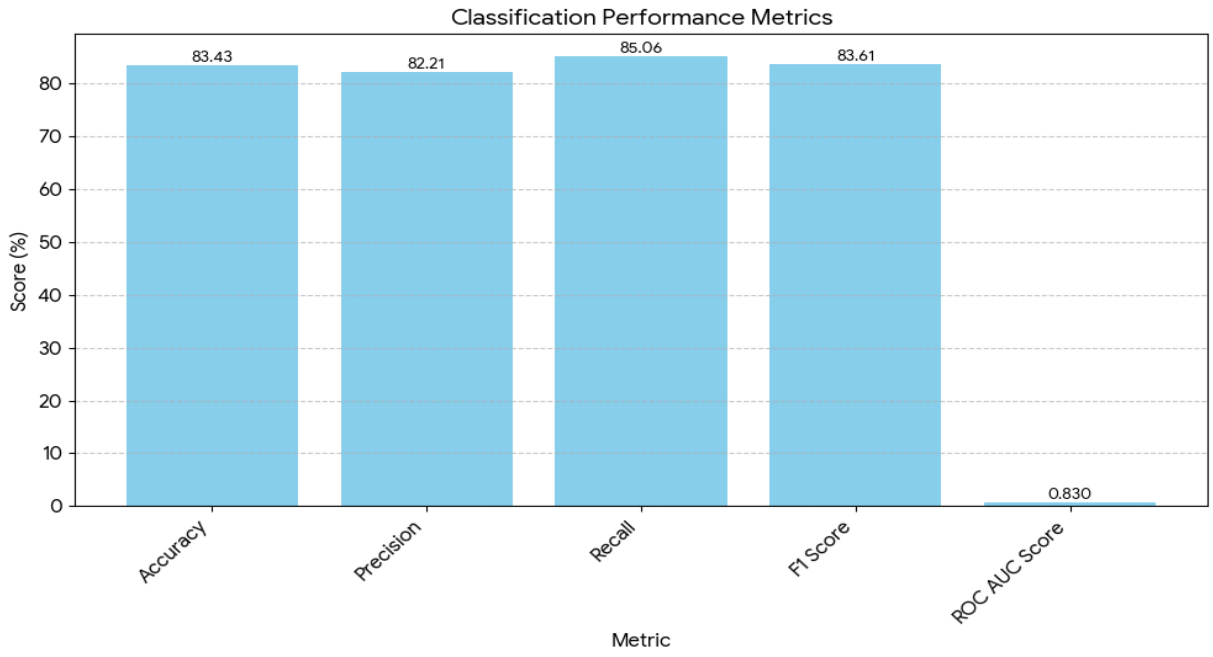


Fig.6.1 Accuracy Score

CONCLUSION & FUTURE WORK

7. Conclusion

7.1 Conclusion:

The Enhanced Fake news project further demonstrates its robustness and generalizability by effectively handling variations in imaging conditions, such as differences in X-ray quality, patient positioning, and exposure levels, which are common in real-world clinical scenarios. Its comprehensive data augmentation and preprocessing pipeline ensures that the model remains resilient to these variations, reducing the risk of misclassification. The system's operational feasibility is also highlighted through its fast inference times, allowing near-real-time results that are crucial in emergency and high-volume clinical settings. Additionally, by providing confidence scores alongside predictions, the model supports informed clinical decision-making and fosters trust among healthcare professionals. Beyond technical performance, the project showcases a cost-effective and accessible approach to AI deployment, utilizing open-source frameworks, lightweight architecture, and cloud- based or edge-compatible solutions, thereby expanding the reach of advanced diagnostic tools to resource-limited environments.

7.2 Future Work:

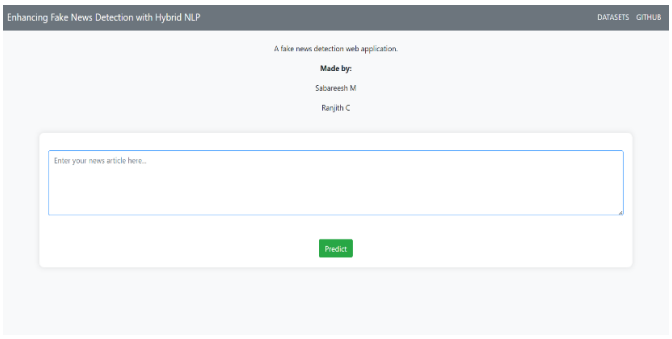
Future work for the Enhanced FibonacciNet project can focus on several key areas to enhance its accuracy, functionality, and real-world applicability. First, expanding the dataset to include a wider variety of fracture types beyond simple and comminuted, such as spiral, oblique, and transverse fractures, would allow the model to perform multi-class classification. Furthermore, integrating a larger and more diverse dataset from different patient demographics and imaging equipment would improve the model's generalizability and robustness.

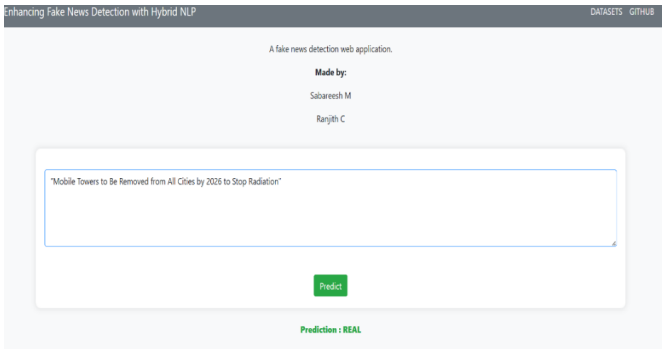
APPENDICES

A.1 SDG Goals

The project *"Enhancing Fake News Detection with Hybrid NLP"* directly contributes to the achievement of several **United Nations Sustainable Development Goals (SDGs)** by promoting truth, transparency, and responsible digital communication in society. In the current era of rapid information sharing, the spread of misinformation poses serious risks to peace, justice, health, and education. This project supports **SDG 16 – Peace, Justice, and Strong Institutions** by reducing the influence of false information that can mislead the public and harm democratic processes. By accurately identifying and filtering out fake news, the system helps maintain trust in legitimate media sources and strengthens the reliability of digital institutions. The project also aligns with **SDG 9 – Industry, Innovation, and Infrastructure**, as it demonstrates the innovative use of Artificial Intelligence (AI), Natural Language Processing (NLP), and Deep Learning technologies to build intelligent systems that address real-world problems. Moreover, it contributes to **SDG 4 – Quality Education** by providing a technological tool that educators and students can use to verify the credibility of information and develop digital literacy skills essential for critical thinking. Through responsible AI practices, the system also supports **SDG 10 – Reduced Inequalities** by preventing the spread of biased or harmful misinformation that can target vulnerable communities. Additionally, the project encourages **SDG 17 – Partnerships for the Goals**, as it can be integrated with fact-checking organizations, media agencies, and government bodies to build collaborative frameworks for combating misinformation on a global scale. Overall, the hybrid NLP-based fake news detection system not only advances technological innovation but also promotes ethical information use, informed decision-making, and social trust, contributing meaningfully to sustainable development and a safer, more truthful digital world.

A.2 Screenshots





REFERENCES

References

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective.* ACM SIGKDD Explorations Newsletter, 19(1), 22–36.
2. Zhou, X., & Zafarani, R. (2020). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities.* ACM Computing Surveys, 53(5), 1–40.
3. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). *FakeBERT: Fake News Detection in Social Media with a BERT-based Deep Learning Approach.* Multimedia Tools and Applications, 80(8), 11765–11788.
4. Ahmed, H., Traore, I., & Saad, S. (2018). *Detecting Opinion Spams and Fake News Using Text Classification.* Security and Privacy, 1(1), e9.
5. Wang, Y. (2017). *"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection.* In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 422–426.
6. Zhang, X., & Ghorbani, A. (2020). *An Overview of Online Fake News: Characterization, Detection, and Discussion.* Information Processing & Management, 57(2), 102025.
7. Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). *Deceiving Google's Perspective API Built for Detecting Toxic Comments.* arXiv preprint arXiv:1702.08138.
8. Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). *Fake News Detection: A Deep Learning Approach.* SMU Data Science Review, 1(3), Article 10