Novel Cognitive Blockchain Based Framework for Mining Electronic Health Records with Secure Patient Centric Communication

L. Prasika[1], T. Ponkumaran[2], N. Sakthi Parameswaran[3]
[1] *Assistant Professor (Sr.Grade), Department of Artificial Intelligence and Data Science Mepco Schlenk Engineering College, Sivakasi, TamilNadu,Inida.*
[2,3] *Student, Department of Artificial Intelligence and Data Science Mepco Schlenk Engineering College, Sivakasi, TamilNadu,Inida.*
doi.org/10.64643/IJIRTV12I12-202275-459

Abstract—**In the present era, the imperative of digitizing of data has become unavoidable. An extensive volume of unstructured Electronic Health Record (EHR) data remains dormant across the internet. Primarily, this data was housed using conventional storage methods. Despite the advantages offered by traditional storage approaches, they frequently grappled with issues such as data loss, absence of immutability concerning health records, and intricate processes for accessing clinical information. Blockchain, a decentralized peer-to-peer storage mechanism, addressed these flaws. EHR data, comprising rich patient-centric information, offer valuable insights for analytical purposes. However, processing this unstructured data and utilizing it for analysis and prediction poses a serious issue. In this research, a novel patient-centric architecture for decentralized health care management with ethereum blockchain using smart contracts and a novel deep neural architecture to extract the features from the clinical nursing notes and prediction of ICD codes is proposed. In this paper, MIMIC IV database was used for training the models and evaluating them. Clinical BERT model was used for converting words into embeddings provides a significant result in prediction. The results obtained from the proposed approach demonstrate a higher accuracy rate of 93% in comparison to alternative models for disease prediction.**

Index Terms—**Blockchain EHR MIMIC Clinical BERT ICD 10 Code Prediction Ensemble Model**

I. INTRODUCTION

Amidst growing interest in self-diagnosis within modern society, extensive research in worldwide has been dedicated to crafting an automated self-diagnosis model capable of predicting diseases based on patient clinical notes as input. A huge amount of unused, unstructured text data is accessible on the internet. Whenever a patient visits a hospital, the hospital takes every piece of information like discharge summaries, the prescription of the doctor etc., in the digital form called as Electronic Health Records (EHRs) unstructured data which has huge amounts of data. The gathered data is stored in the digital form as EHR. The EHR contains details regarding individuals' visits to doctors or hospitalizations, encompassing information such as past medical history, family medical history and more. Most healthcare institutions predominantly employ centralized databases for storing such information with online access posing various challenges including interoperability, privacy, data security and scalability issues. In the initial phases,

requirement of data sharing and access in real-time,healthcare institutes have turned to cloud-based data management, necessitating encryption and decryption of Distributed storage systems were employed to tackle these challenges yet distributing EHRs proved time-consuming

and financially burdensome. Subsequently, to address the data for each transfer to and from the cloud thereby rendering it a resource-intensive process. To tackle these challenges, we introduced a blockchain-based data management system which is used to store and maintain EHRs. Blockchain, a distributed, decentralized, and peer-to-peer digital ledger, is utilized to record and store transactions, which become immutable once stored. Employing cryptographic methods, blockchain encrypts stored data using both private and public keys, thereby restricting access solely to authorized users. This characteristic enables blockchain to furnish a dependable communication platform facilitating secure information transactions among peers. Blockchain also resolves issues such as record privacy, access control, confidentiality and data integrity. The collection and preservation of patient data in digital format emphasized the significance of early disease diagnosis. Thus far, the health care industry is limited to the diagnostic and actions to be taken for the diagnosis. With the advancement in technologies, these barriers must be removed. The Clinical Decision System (CDS) was developed with the help of the MIMIC IV database which consists of various health records and other related data in the colossal amount. CDS has improved prediction skills, which are critical for promoting patient-centered and evidence-based treatments, lowering morbidity and mortality rates, and facilitating better risk assessment. Unstructured data in EHRs are tagged and contain important health information such as symptoms, procedures, diagnosis codes, prescriptions, surgeries, and test results. This unstructured nursing notes contain various information necessary for diagnosing the patient. Our approach includes:

- Training the proposed model by leveraging the MIMIC IV database.
- Subsequently, identifying the corresponding ICD-10 code for each disease.
- The Clinical Bio BERT model was used in converting words to embeddings which provide the efficient result than compared to the other models.
- Employing the Ethereum Blockchain to facilitate secure communication among the entities and ensuring secure storage of data through utilization of the Ethereum Blockchain.

II RELATED WORKS

With an exploration of diverse methodologies and advancements in healthcare, the literature study delves into the complex structures of EHRs. The employment of the Ethereum Blockchain a decentralized ledger technology, to fortify communication channels among various entities within the healthcare domain. The survey looks at how important blockchain technology is to guarantee the safe keeping of private medical records. Following this introduction, next discusses the specifics of the research paper, elucidating its innovative approach towards decentralized health data management and the utilization of deep neural architectures for predictive analytics in healthcare. Through a meticulous analysis of pertinent studies and methodologies, the paper endeavors to contribute novel insights into the intersection of blockchain technology, deep learning and healthcare management.

2.1 Analytical Methods For Uncovering Patterns in Clinical Narratives

In [1], Tushaar Gangavarapu developed deep learning models for predicting disease based on clinical notes, introducing a NLP pipeline for preprocessing. The study employed various models such as Convolution LSTM and Bidirectional LSTM showcasing promising results in disease prediction. Xiaowen Zhang et al. [2] provided a description of DL techniques in healthcare, highlighting their applications and challenges. The study discussed the potential of DL models in medical image analysis, EHR analysis, and genomics, emphasizing the need for addressing data heterogeneity and privacy concerns. In [3] compared the performance of Convolutional Neural Networks (CNNs) with BioBERT in clinical text classification. The study suggested that CNNs might outperform BioBERT due to complexities in BERT's pretrained process and tokenization. [4] introduced Med-BERT, a Named Entity Recognition (NER) with BERT model in medical records. The study proposed a novel loss function and demonstrated superior performance, particularly in recognizing long medical entities. [5] emphasized the analysis of unstructured medical text data using Natural Language Processing (NLP) tools.It highlighted the importance of transformer-based models and effective data preprocessing for accurate outcome prediction.

Minji Kim et al. [6] examined the use of deep learning methods for patient symptom-based medical diagnosis.The study introduced Symptom2Vec, which learns relationships between symptoms and diseases, facilitating automatic diagnosis. [7] proposed a deep learning approach for using features from medical knowledge in clinical event extraction to improve model performance. This study highlighted the potential of domain knowledge in improving clinical text analysis. [8] explored the use of deep learning models for extracting symptoms from the clinical text and social media data focusing on COVID-19 symptoms. The study discussed the challenges of analyzing informal social media data and potential biases.[9] investigated the implementing Clinical BERT model with models for predicting multimodal mortality for COVID-19 patients. The study explored different strategies for leveraging Clinical BERT and highlighted the importance of high-quality clinical data for model training. [10] proposed using domain expertise to improve clinical text Natural Language Inference (NLI) tasks. The study discussed the potential benefits and challenges of integrating medical ontologies into NLP models. [11] presented an accurate deep learning model for identifying identifying clinical entities from clinical notes using clinical entity recognition. The study emphasized the potential of deep learning models in automating the extraction of clinical information from textual data. [12] explored the use of gated attention-BiRNN for identifying heart failure based on clinical notes analysis. The study emphasized the limitations of training data biases and the significance of including medical knowledge elements. [13] examined the intersection of clinical big data and deep learning, discussing its potential, obstacles, and uses in healthcare. The study emphasized the importance of ensuring patient privacy and understanding model reasoning in high-stakes medical applications. [14] applied deep learning techniques to analyze EHRs, discussing its potential in disease prediction and treatment optimization. The study highlighted challenges such as data inconsistencies and privacy concerns. [15] provided an overview of Deep Neural Networks (DNNs) for ICD coding, emphasizing their potential in automating the coding process and improving efficiency in healthcare coding tasks. The study discussed various DNN architectures and their applications.

Saddam et al. [16] investigated the application of machine learning for classifying patients at risk of opioid misuse. The study discussed the potential benefits of early identification of at-risk patients and challenges related to training data quality. [17] utilized machine learning to predict the number of days ICU patients stay by using the MIMIC-III database. The study discussed the potential benefits of accurate length-of-stay prediction in optimizing ICU resource management. [18] explored a deep learning method for classifying diseases using NLP techniques and MIMIC-III database.The study discussed the potential benefits of automating disease classification from clinical notes. [19] focused on using deep learning for data filtering within the MIMIC-III dataset to enhance reliability and generalizability of research findings. The study emphasized the importance of automating data filtering processes. [20] compares the MIMIC databases,both MIMIC-III and MIMIC-IV for gathering health informatics for big data analytics , discussing their data size, structure, quality, and use cases. The study highlighted the potential benefits and limitations of leveraging both datasets for healthcare research.[21] proposes a DL based Named Entity Recognition model which combines Bi-LSTM architecture with conditional Random Field for addressing the detection of product names from the unstructured Turkish text.The NER with Bi-LSTM and CRF architecture have seen a significant advancement by using the pretrained embeddings and neural network architecture.[22] proposes a deep neural architecture using the combination of pretrained BERT model and layers of Convolutional neural networks(CNN). The BERT model provides the encoding of the text data to numerical data which is used to capture the meaning of words.while CNN helps to identify the patterns from the encoded text data related to suicide risk.This improves the perfomances of the model in detecting suicide risk.

2.2. Blockchain-Enabled Secure Communication

Akhilendra Pratap Singh et al. [22] provide a platform for blockchain-enabled healthcare applications that are patient-centric communication,addressing data privacy, authentication, and deployment strategies using the Hyperledger platform. [23] focus on securing patient information sharing through blockchain-based distributed networks, emphasizing privacy concerns and patient consent management. [24] highlight the need for secure and privacy-preserving methods in disease management with less emphasis on scalability and computational costs. [25] discuss blockchain's potential in healthcare, emphasizing data security and interoperability while overlooking scalability and regulatory hurdles. [26] propose a patient-centric access control system using dual-blockchain architecture, prioritizing fine-grained access control and scalability concerns. [27] develops a medical data exchange plan for e-healthcare systems that protects privacy, focusing on data anonymization and cryptographic techniques.

Ramzan et al. [28] explore blockchain applications in healthcare, stressing immutability and decentralization for secure data exchange. [29] present a blockchain-based encryption scheme for EHRs, ensuring the integrity of data and access control. [30] delve into processing health insurance claims using blockchain highlighting benefits such as data integrity and automation, while addressing challenges like regulatory compliance. These studies collectively reflect the growing interest in leveraging blockchain technology to address various challenges in healthcare, while also acknowledging the need for further research on scalability, regulatory compliance, and integration with existing systems.[31] propose A novel approach to use block compression and IPFS to increase Bitcoin's scalability, addressing the storage demands on individual nodes within the Bitcoin network. [32] introduce PDPChain, A distributed blockchain-based privacy protection program for publishers of personal data, highlighting improved data security via on-chain transmission and off-chain storage. [33] present a blockchain and smart contract system for social security services offering increased transparency and efficiency through automated processes and tamper-proof data storage. [34] propose a decentralized model leveraging utilizing the Polygon blockchain's smart contracts to safeguard digital evidence aiming for addressing vulnerabilities in centralized systems susceptible to data manipulation. [35] introduce Artificial Identification, a blockchain-powered privacy architecture for federated learning aiming to protect data privacy while enabling collaborative model training. [36] propose a data sharing program that combines attribute-based encryption (ABE) with blockchain, emphasizing enhanced data privacy through anonymity and attribute-based access control. Mikail Mohammed Salim et al. [37] propose a secure Electronic Health Record (EHR) sharing program in medical informatics that is built on federated learning with the goal of improving healthcare insights while preserving patient privacy through data encryption.

After the comprehensive review of the literature study, we infer that the EHRs role in the healthcare domain is inevitable. Bio Clinical BERT delivered high-quality results in case of EHRs as it even takes the small and important information between the words in the clinical notes. The use of Blockchain in communication is also promising and secure.

III PROPOSED METHODOLOGY

3.1 Dataset and Specifications

The MIMIC-IV dataset stands as a comprehensive and widely utilized repository containing de-identified health-related data from patients admitted to critical care units at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, USA. Spanning a multitude of years and encompassing tens of thousands of patients, this dataset offers a rich array of clinical data, such as patient demographics, vital signs, lab results, prescriptions, treatments, diagnoses, and clinical notes both in structured relational database format and unstructured free-text format. Researchers and healthcare professionals leverage the MIMIC-IV dataset for a myriad of purposes ranging from clinical research and predictive modeling to algorithm development and healthcare analytics. Its extensive usage has significantly advanced research in areas such as predictive modeling for patient outcomes, clinical decision support systems, disease progression analysis and understanding critical illness epidemiology. Access to the dataset is regulated by the MIT Laboratory for Computational Physiology requiring approval through a formal application process to ensure compliance with privacy and ethical considerations. Researchers typically adhere to stringent guidelines and sign data use agreements to maintain patient privacy and data security. The dataset undergoes periodic updates to incorporate additional data and improvements based on feedback from the research community cementing its position as a vital resource for advancing clinical research and MIMIC IV has various modules like hospital, icu, electronic deidentified text(ed), cxr and note. In this, the hospital module contains various hospital wide EHR. The icu module contains ICU level data like chart events and other even tables. The ed module contains only the data related to the emergency department. The CXR module contains chest x-ray images. The note module contains the discharge summary of all the patients. The key update from MIMIC III to MIMIC IV involved a complete modularization of the dataset along with the inclusion of additional data. The information presented in the notes is comprehensively detailed in Table 1. The Chief Complaint delineates the patients' reported issues to the

doctor, whereas a Major Surgical or Invasive Procedure refers to the surgery that has been carried out, History of Present Illness goes beyond a mere summary offering meticulous description of the primary complaint, past medical history encapsulates the patient's health issues that the patient had in the past, family history highlights the genetic issues within the family, physical exam encompasses a thorough evaluation about the Vital Signs such as temperature, heart rate, blood pressure, breathing rate, and oxygen saturation. General observation assesses the patient's overall appearance and general health status and also includes observations about distress, signs of illness or any other noticeable abnormalities, Neck, and Description

HEENT (Head, Eyes, Ear, Nose, Throat), CV (Cardiovascular), Lungs, Abdomen, GU (GenitoUrinary), Ext(Extremities) and Neurological. HEENT involves a thorough evaluation of the head and its primary sensory organs. This examination encompasses a close examination of the eyes, ears, nose and throat to detect issues pertaining to vision, hearing and upper respiratory functions. Neck examination assesses the range of motion, presence of lymph nodes (lymphadenopathy) and any other abnormalities in the neck area. By scrutinizing these key areas, professionals can easily detect and address potential problems related to sensory and respiratory perception. CV pertains to examinations that focus on the heart and blood vessels. This involves the heart sounds checking for the abnormal rhythms or murmurs and evaluating the pulse and blood pressure. Lung examination entails listening to the sounds produced during breathing. Abnormalities like wheezing, rales or decreased breath sounds indicate respiratory issues, fluid accumulation or organ-related problems. This thorough evaluation of the cardiovascular and respiratory systems is crucial for a comprehensive understanding of a patient's health. GU directs attention to an examination of the genital urinary organs. This comprehensive assessment assesses the presence of catheters, signs of inflammation or infection and any abnormalities in the genital or urinary tract. Ext pertaining examination evaluates the arm and legs for signs of swelling deformities joint abnormalities or circulation problems. This comprehensive assessment includes assessing muscle strength and tone. Neuro is the Neurological assessment of the patient's motor and sensory capacities, reflex coordination, cranial nerve functions, muscle status, and gait.

Table.1 MIMCIV Features

Attribute	Description
Chief Complaint	The complaint or the symptoms that the patient face and it is given by the patient itself.
Major Surgical or Invasive Procedure	The surgery that has been carried out.
History of Present Illness	Detailed description of Chief Complaint.
Past Medical History	Disease that patient had before.
Family History	Tells the Genetic Issues
Physical Exam	Describes VS, General, HEENT, Neck, CV, Lungs, Abdomen, GU, Ext, Neuro.
Discharge Diagnosis	Diagnosed health issues.

IV. SYSTEM DESIGN

The proposed system contains two modules Blockchain Framework and EHR Mining Framework (EMF). This Blockchain Framework consists of three entities Patien, Doctor, Chemist followed by the EMF contain the following modules Feature Extraction, Dictionarization, Feature Modeling, Dimensionality

Reduction and the ICD Code Prediction using proposed model. The ICD code predicted from the EMF is communicated securely between the entities over the blockchain. In this research, Ethereum blockchain serves as the foundational technology for the study. Ethereum is a decentralized blockchain platform renowned for supporting smart contracts and decentralized applications. By utilizing Ethereum's capabilities, this research aims to leverage smart contracts and decentralized systems to address various aspects relevant to the research. Ethereum's features including its support for smart contracts, decentralized execution and native cryptocurrency Ether (ETH) provide the essential infrastructure for implementing innovative solutions. EHRs represent an electronic version of a patient's whole medical history that includes a wealth of data, including diagnosis, prescriptions, plans of care, dates of vaccinations, allergies, radiological pictures, lab test results, and more. They serve as a central database containing patient health information that authorized healthcare practitioners can access across different healthcare settings. EHRs facilitate seamless communication and information sharing among healthcare professionals enhancing coordination of care and improving patient outcomes.

EHRs provide a number of benefits over traditional paper-based systems by digitizing health records, including increased efficiency, accuracy, accessibility and security. Additionally, EHR systems often incorporate features such as decision support tools, reminders and alerts to assist healthcare providers in delivering evidence-based care and reducing medical errors. Figure 1 describe the overall architecture of the proposed approach, which encompasses the secure communication among entities and HER mining framework.

4.1 Blockchain Entities

Entities were participants. Participants can be able to transfer assets between the other participants. In this application, there are three participants. They were patients, doctors and chemists.

4.1.1. Patients

Every patient has a unique patient id in this application, patient address becomes their patient id since it is unique and whenever a patient visit the hospital,he/she will be provided with a hospital admission id.The patients had functionalities for predicting the disease by providing clinical notes.The clinical notes are provided to the EHR mining framework which predicts the ICD codes for given symptoms and store it securely in the Ethereum blockchain. Patients can fix an appointment with the specialist if any time slots are available. They can also access their medical details as well as prescriptions from the blockchain ledger either through their hospital admission id or patient id that is their account address.

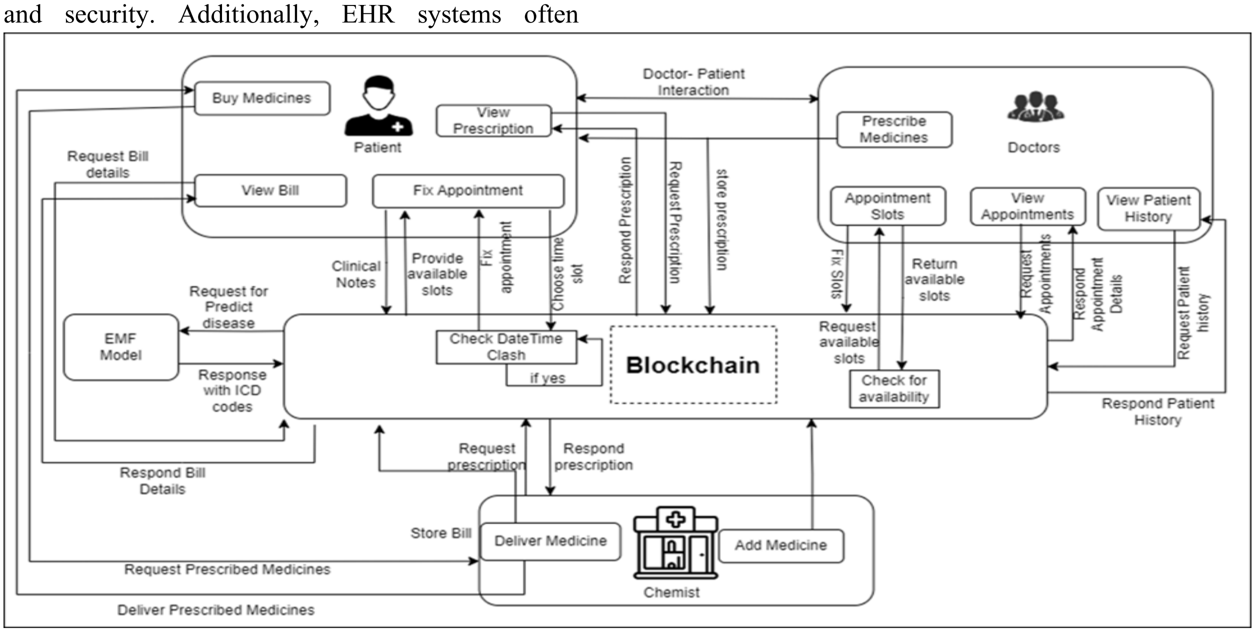


Fig. 1 Workflow architecture

4.1.2 Doctors

Doctors have functionalities like fixing appointment slots which is the number of patients, they attend on that day. Based on this, the available slots for the patients will be displayed. The doctors can provide a prescription for each patient and can be able to view the patients' medical history and can view and edit the appointments. The prescriptions to the patient were stored on a blockchain which can then be accessed by the corresponding patient as well as the chemist.

4.1.3 Chemists

Chemists had functionalities such as maintaining medicine details in the database and delivering medicine to the patient. In delivering medicine to the patients, the patients send a request to the chemist, then the chemist retrieves the prescription of the particular patient from the blockchain network. Then the medicines will be delivered to the patients once the bill is paid. The transactions happen in "wei", a smallest denomination of ethers.

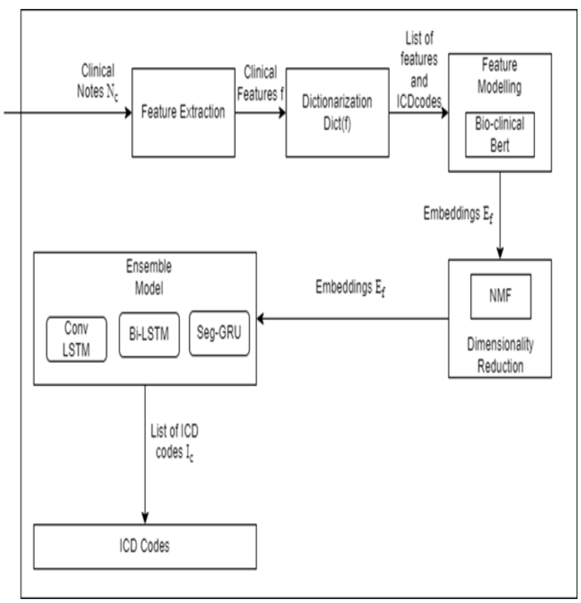


Fig.2 EHR Mining Framework

4.2 EHR Mining Framework

The architecture for the deep learning model was designed for predicting ICD codes from clinical notes begins with a series of preprocessing steps aimed at optimizing the input data. Firstly, the clinical notes were subjected to feature extraction where relevant information was identified and extracted from the text. Subsequently, the extracted features undergo dictionarization a process wherein the identified terms were mapped to a standardized vocabulary ensuring consistency and uniformity in representation. Following dictionarization, the features were structured and organized through Feature Modeling, which involves arranging them into a suitable format for input into the deep learning model. Finally, Dimensionality reduction techniques may be applied to streamline the dataset, reducing its complexity and improving computational efficiency while retaining crucial information essential for accurate prediction. Through these preprocessing steps, the architecture prepares the clinical notes data for effective utilization by the deep learning model, ultimately enhancing the accuracy and reliability of the ICD code predictions.

4.2.1 Feature Extraction

Within the MIMIC-IV dataset, three distinct modules were available: hospital, ICU, and deidentified text notes (DT) along with additional data on chest X-rays (CXRs). The DT module encompasses a discharge table, housing discharge summaries of patients, characterized by attributes such as note_id, subject_id, hospital_admission_id, note_type, and text. Specifically, the text attribute within the discharge table encapsulates the comprehensive discharge summaries of patients. From this textual data, six key features were extracted, including information on allergies, chief complaints, major surgical or invasive procedures, history of present illness, family history and past medical history. These features serve as an essential components in analyzing patient health records and contribute to the comprehensive understanding of their medical history within the dataset.

4.2.2 **Dictionarization**

The hospital module within the MIMIC-IV dataset comprises 22 tables. Specifically, the diagnoses_icd table, a component of this module includes attributes such as subject_id, hospital_admission_id, seq_number, icd_code, and icd_version. To facilitate comprehensive analysis, the discharge table from the emergency department (ED) module was aggregated with the diagnoses_icd table from the hospital module. This aggregation was based on the common

$O\alpha^T$

attribute of hospital_admission_id present in both tables. By merging these tables using the hospital_admission_id as the key identifier, a holistic view of patient care and outcomes across different healthcare settings within the MIMIC-IV dataset is achieved.

4.2.3 Clinical Feature Modeling

Clinical feature modeling refers to the process of identifying, analyzing and representing the specific characteristics or features in clinical data to gain insights into disease, treatments and patient outcomes. In the context of healthcare and medical research, clinical features can encompass a broad variety of factors, including the patient's medical history, vital signs, vital signs, test results, and symptoms. But these features cannot be provided to the model in its existing textual format. These features must be transformed to the word vector embeddings. Word vector embeddings, sometimes called word embeddings, are numerical representations of words in continuous vector space that are arranged next to one another according to similar contexts or meanings. These embeddings are generated using Clinical BERT.

4.2.3.1 Clinical BERT

BERT (Bidirectional Encoder Representaions from Transformers) is a transformer-based neural network model designed for challenges involving natural language processing. The Clinical Bert model was fine-tuned using a large-scale corpus of EHRs from over 3 million patient records after being trained on a huge multicenter dataset like MIMIC, which contains a big corpus of 1.2 billion distinct diseases. It was unique because it reads text bidirectionally and can express a sentence's meaning for each word dependent on the ones that surround it. The core concept behind the BERT and transformer architecture was the self-attention (SA) mechanism. Self-Attention enables the model to balance the significance of each word in a sentence when encoding the representations of those words, which allows BERT to capture contextual information effectively. The SA score for a pair of words is calculated using equation 1.

$$\text{Attention}(O, \alpha, \lambda) = \text{softmax}\left(\frac{O\alpha^T}{\sqrt{A_b}}\right)\lambda \qquad (1)$$

Where:

O represents the query matrix

α represents the key matrix

λ represents the value matrix

A_b represents the dimensionality of the key vectors.

$$\left(\frac{O\alpha^T}{\sqrt{A_b}}\right)$$

is the scaled dot-product attention, where represents query matrix and key matrix dot product and the division of the dot products by scales.

BERT employs multi-head self-attention, in which the attention mechanism is used again in parallel, each time using a different set of input embeddings that have been learnt. The final output embeddings for each subsequent layer in the model are obtained by concatenating and linearly transforming the outputs from these parallel attention heads. Furthermore, BERT presents the idea of "masked language modelling," in which a subset of the input sequence's words are hidden, and the model is trained to predict these hidden words by utilizing the context that the other words in the sentence give. BERT’s training objective involves maximizing the likelihood of predicting masked words in a sentence, both from the left and right contexts, across multiple layers of the model. Because of its ability to learn rich contextual representations through bidirectional training, BERT is incredibly efficient for a variety of natural language processing applications. A modified form of the BERT model called Clinical Bert was created especially for handling text data related to medicine and healthcare. Unlike traditional BERT, Clinical BERT was trained on large dataset containing medical text, allowing it to understand complex medical terminology abbreviations, and context-specific language patterns. By incorporating domain-specific vocabularies and training techniques, Clinical BERT excels in capturing and nuances of clinical languages, making it invaluable for various healthcare applications, from electronic health record analysis to clinical decision support systems. A variation of standard language modeling called masked language modeling was used to train clinical BERT. When using masked language modeling, a model is trained to predict words that are masked in input sequences by using the context that the other words in the sentence give. According to naives aggregate the clinical notes based on the patient identification number subject_id.Let the notes be the $\{t_1^{(p)}, t_2^{(p)}, \ldots.. t_n^{(p)}\}$on aggregating we get the ICD code as the following $\{k_1^{(p)}, k_2^{(p)}, \ldots.. k_n^{(p)}\}$. On mapping the clinical notes to the ICD codes, we get the ICD code as $\{k_1^{(p)} + k_2^{(p)} + \ldots.. + k_n^{(p)}\}$.Here, p represents the patient identification number.For prediction, calculate the mean of the embedding $\{e_1^{t_1^{(p)}}, e_2^{t_2^{(p)}}, \ldots, e_n^{t_n^{(p)}}\}$.

4.2.4 Dimensionality Reduction

NMF or Non-negative Matrix Factorization, was a technique used in machine learning for dimensionality reduction and feature extraction. The goal of NMF was to factorize a given matrix into two non-negative matrices, whose product approximates the original matrix, NMF was particularly useful when dealing with the non-negative data such images, text documents or other types of non-negative matrices. Let's represent the original matrix as V, which is of size m x n, and we want to factorize it into two non-negative matrices ώ and H where ώ has dimensions m x n and H has dimensions r x n. Here, r is desired rank or the number of components. It can be formulated as in equation 2.

$$V \approx ώ.H \quad (2)$$

Here, ώ and H were subjected to the constraints that all the elements must be non-negative in other words, $ώ_{ij} \geq 0$ and $H_{ij} \geq 0$ for all i and j. The objective function to minimize the approximation error can be defined using various distance metrics, with the Euclidean distance being the common choice. The objective function of NMF with Euclidean distance can be written as in equation 3.

$$\text{minimize } ||V - ώH||_F^2 \ldots\ldots\ldots\ldots(3)$$

Here, $|\cdot|_F$ denotes the Frobenius norm, which was generalization of the Euclidean norm for the matrices.

4.2.5 Ensemble Model

Model Ensembling helps in the universality of the model. It can simply define as the technique of combining the multiple models in the diverse ways to improve the performance on a single problem. Since clinical notes contains different types of information like chief complaint, present illness, past illness, family history etc. Each model in the ensemble model excels at capturing different aspects of this information. Like, Conv-LSTM might adept at identifying sequential patterns in the clinical notes, Bi-LSTM excels at capturing bidirectional dependencies and Seg-GRU effective at segment-level analysis. On concatenation of the Seg. GRU, Conv-LSTM and BILSTM model, we create the ensemble model. So, by combining these models, the proposed ensemble model can effectively capture and integrate different types of data present in the clinical notes.

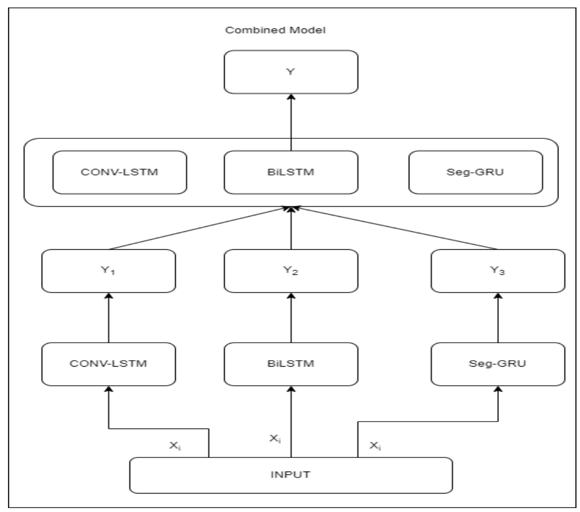


Fig.3 Ensemble Model Architecture

4.2.6 Convolutional Long Short Term Memory architecture

The convolutional layer effectively extracts the high-level features from the clinical nursing notes. To capture the long-term dependencies in the nursing notes embed it with the convolutional network. Clinical notes were sequential in nature, each word or sentence had contextual relevance to the other word or phrase either it may be one before or after. LSTM network can retain the long-term dependencies, this makes LSTM network well-suited for processing sequential data. The convolutional layer can effectively extract high level features from the input data. In clinical notes, these features represent linguistic pattern syntactic structure, semantic meaning, etc. By using the Convolution layer, the model can learn meaningful representations from the raw text data. It has nine layers, including the input and the output layer. It has one convolutional layer with 19 filters and kernel size of 4 and activation function used here is ReLu. The Dense layer is the Fully connected layer with the L2 Regularization with 289 neurons and activation function of ReLu and regularization of L2. The dropout layer is used for dropping the layer which is less important for ease computation. The Dropout rate specifies that the certain percentage of the neurons will be randomly dropped out during training, which helps to prevent overfitting. There is one more convolutional layer with 19 filters used and kernel size of 3. Then the Dense layer is incorporated and following this LSTM layer is embedded with the units of

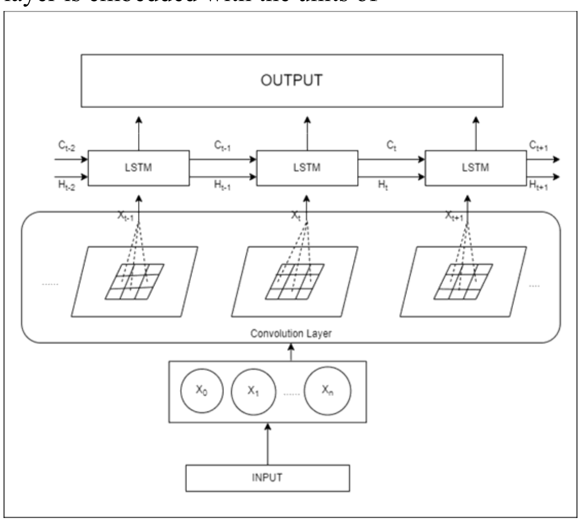


Fig.4 Conv-LSTM Architecture

300 LSTM units. Output layer is used with the activation

function called sigmoid. So, using the proposed Conv-LSTM model allows effective modelling of sequential data and capturing of long-term dependencies which improves prediction accuracy.

4.2.7 Bi-LSTM

Bidirectional Long Short-Term Memory is a recurrent neural network (RNN) architecture that improves the capabilities of regular LSTM networks by processing input data in both forward and backward directions. Bi-LSTM has grown in popularity in the fields of natural language processing and sequential data modeling due to its capacity to gather contextual information from both past and future tokens in a sequence at the same time. Understanding the context surrounding a particular word or phrase and the relationship between different parts of the clinical notes, Bi-LSTM was used in this approach. Two LSTM layers make up Bi-LSTM: a forward LSTM processes the input sequence from start to finish (forward LSTM), while a backward LSTM processes the sequence from start to finish (backward LSTM). By doing this, the model is able to comprehend the input sequence more thoroughly by capturing patterns and dependencies from the past as well as the future context of each time step.To capture bidirectional contextual information, learn informative representations, handle long-term dependencies and accommodate input variability, Bi-LSTM is used

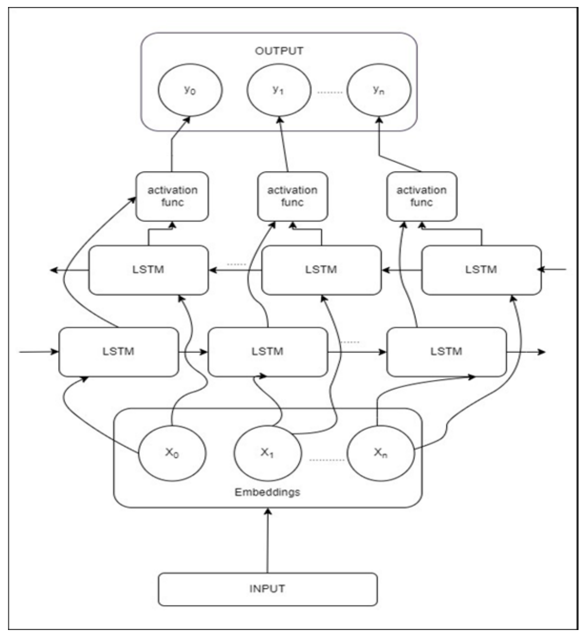


Fig.5 Bi-LSTM Architecture

4.2.8 Seg -GRU

Segment Level Gated Recurrent unit is part of GRU. The GRU is the one of the gating mechanisms in RNN, which is similar to the LSTM. The GRU is made up of two gates: the reset gate (re) and the update gate (up). It can be formulated from the equation (4) - (6)

$$\binom{re}{up} = \binom{sigm}{sigm} p^{(l)} \binom{k_t^{(l-1)}}{k_{t-1}^{(l)}} \quad ....(4)$$

$$\widehat{k_t^{(l)}} = tanh(w^{(l)}.k_t^{(i-1)} + I^i(re \boxdot k_{t-1}^i)) \quad ....(5)$$

$$k_t^{(i)} = (1 - up) \boxdot k_{t-1}^{(i)} + \mathrm{up} \boxdot \widehat{k_t^{(i)}} \quad \ldots\ldots(6)$$

Where $\boxdot$ denotes the Hadmarad product, $p^{(i)}$, $w^{(i)}$ and I^i are the weight matrices at the layer (l) and $k_t^{(i)}$ is the hidden state at the time step t and the candidate hidden state $\widehat{k_t^{(i)}}$ is computed using GRU. In the Segment level gated recurrent unit, Following the column-wise splitting of the input embeddings into 20 segments, the fully connected layer of Q+Q/20 ReLu processing units is placed next, where the Q represent the nursing Notes, and the other fully connected layer of 20.Gated Recurrent unit have a simpler architecture compared to LSTMs, which makes faster training times and fewer parameter to tune.

V. Algorithmic Framework

Table 2 Algorithm Notations and Description

Symbols	Connotation
N_c	Clinical Notes
I_c	ICD diagnosis
ψ	Mapped notes with corresponding ICD codes
F= [f1, f2.... fn]	F represents the features where f1, f2, fn represents features like chief complaint,major surgical,allergies,etc,.
Dict(.)	Dictionarizing Function
E	Embedding
V	NMF vector value
ConvLSTM(.)	Convolution LSTM
BiLSTM(.)	Bidirectional LSTM
SegGRU(.)	Segment Level Gated Recurrent Unit
Ensemble(.)	Ensembled Model
Δ	Address which login in
$Doctor_{node}$, $Chemist_{node}$, $Patient_{node}$	Addresses of the doctor,chemist,patient.
ϐ	List of specialization
ʝ, ꝣ, ƨ, ƻ	Date, Start time, End time, Max Patients
S_{id}	Appointment Slot id
C_c	Chief Complaints
F_H	Family History
P_I	Present Illness
P_{MH}	Patient Past Medical History
P_{MS}	Patient Major Surgical
P_a	Patient Allergies
P_{icd}	Icd Codes
hadmn_id	Hospital admission id
ɋ	Drugs
ꝑ	Prescription

The proposed algorithmic framework delineates a comprehensive system for leveraging blockchain technology in healthcare management, focusing on patient care coordination and transactional transparency. This framework integrates deep learning techniques with blockchain-enabled functionalities to streamline processes such as patient enrollment, appointment scheduling, prescription management, and drug dispensation. Through a series of interconnected algorithms, it aims to improve the efficiency, accuracy and security of healthcare operations while ensuring compliance with regulatory standards. By harnessing the power of blockchain for data immutability and transparency coupled with advanced deep learning models for clinical data analysis and prediction the framework offers a promising avenue for revolutionizing healthcare delivery. This research presents an in-depth exploration of the algorithmic framework, detailing its components, functionalities, and potential applications in modern healthcare settings.

Algorithm 1: Preprocessing of clinical notes

Input: Nc (Clinical Notes), I_c (Mapped ICD code for the Patients)
Output: ψ (Mapped notes with the ICD codes)
1. for each ς in N_c:
2. F=Extract F (ς)
3. dictF = Dict(F)
4. ψ += Group dictF and I_c by hadmn_id
5. End for
6. Return ψ

In the above illustrated algorithm 1, the N_c represents number of notes, I_c represents mapped ICD code for the corresponding patients and ψ represents the mapped notes with the ICD codes of the respective patients which was the main object of the fore mentioned algorithm. The Algorithm starts by Processing each notes N_c. Each note ς contains the information about the patients as well as the features like allergies, chief complaint, major surgical, history of present illness, past medical history as shown in the table 1. These features were then extracted using simple regular expressions and stored in the F using the function F(.). After the feature extraction of the each clinical notes, a single note contains maximum of seven features,these features were converted into key value pair like allergies with its information and so far. Using Dict(F) function the features were converted into dictionaries and stored in the dictF variable.This process was termed as dictionarization which was done for every notes.In the dataset, a individual may visit the hospital for more than one time for different reasons which makes the individual patient to have many clinicl notes under same hospital id(hadmn_id) and also have different ICD codes for an individual. So by using the hadmn_id the dictF and I_c are grouped and return ψ to the Algorithm 2.

Algorithm 2: Creating Embedding for ψ

Input: ψ , I_c

Output: V

```
for each ti in ψ:
    for each l in ti:
        for each f in l:
            Ef = BioClinicalBert(f)
        End for
    AE = AVG(Ef)
    End for
E += AE
End for
V= NMF(E)
Return V
```

The pre-processed notes ψ from the algorithm 1 and I_c mapped ICD code were given as the input . For eachnote t_i in ψ has the notes as list of clinical notes which were grouped using hospital admission id (hadmn_id). Each clinical notes from each ψ and form each ψ every note l the features are taken and given as input to the BioClinical Bert model which was pretrained model for converting the clinical notes into embeddings E_f. Each features f from notes l is tranformed into embeddings E_f. After that mean of the features (AE) of each embeddings E_f are found for every note. It was then stored as AE(Average of Embeddings). For dimensionality reduction the embeddings E_f were given as an input to the NMF (Non-Matrix Factorization) function stored and returned to the Algorithm 3 as V

Algorithm 3: Predicting ICD Code I_c

Input: V

Output: I_c

I_1 = ConvLSTM(V)
I_2 = BiLSTM(V)
I_3 = SegGRU(V)
I_4 = Ensemble (I_1, I_2, I_3)
Return I_4

The transformed features V from the algorithm 2 was given as the input, then this features were feed to models like ConvLSTM(.), BiLSTM(.) and SegGRU(.) respectively, then Ensemble all the predictions I_1, I_2, I_3 and return I_4

Algorithm 4: Algorithm on participant enrollment and functionalities.

Input: Δ

Output: Storing transaction.

```
While Δ is valid || Δ is admin do:
    if Doctor node is valid then:
        AddDoctor(Daddr , Dname ,Ϭ);
    else:
        not Valid (Doctor node);
    end
    if Chemist node is valid then:
        AddChemist(Caddr , Cname);
    else:
        not valid (Chemist node);
    end
    if Patient node is valid then:
        AddPatient(Pname ,Page ,Ploc , PGender);
    else:
        not valid (Patient node);
    end
end
if Δ == Doctor node then:
    addAppointmentSlot(ȷ, ꝛ, ƽ, ϱ);
    assignDrugs(Paddr , hadmn_id , p);
end
if Δ == Patient node then:
    fixAppointment(Sid , Daddr , Cc, FH ,PI, PMH , PMS , Pal );
    BuyDrugs(Paddr, hadmn_id);
    GetBill(Paddr, hadmn_id);
end
if Δ == Chemist node then:
    addDrugs(ḑ name, ḑ dose, ḑ price);
    DelieverDrug(Paddr, hadmn_id);
end
```

The algorithm illustrates the transactions that were done by different participants in the blockchain network. Δ denotes the address that the user used to login, if the Δ is admin, he has the authority to add patients, doctors and chemist details. If the address of the patient or doctor or chemist is invalid, then the details will not be added but the transaction for trying to add the wrong details also stored in the ledger which implies the transparency of the blockchain. If the login user is doctor, the doctor has the functionalities for adding appointment slots which takes the input like date ȷ, start time ꝛ, end time ƽ, Max patients ϱ and can be able to assign drugs to each patient. The doctor uses the patient's id (P_{addr}) and hospital admission id (hadmn_id) for accessing each patient and assign prescription. If the user address is patient, the patient can be able fix appointment by providing the details chief complaint, family history, Present Illness, Patient Past Medical History, Patient Major Surgical, Patient Allergies. This data will be provided to the deep learning model which preprocess the data and provide it to the ensemble model which predicts the ICD codes I_c and check the availability of the slot for the specialist and appointment slot will be fixed. Then the hospital admission id (hadmn_id) for the patient will be returned. The patient can get bill for the hospital admission id. If the user address was chemist, the chemist can be able to add drugs with drug name, drug doses and price for the medicine. When the patient request for the medicine, the chemist access the prescription from the network by using the patient address (P_{addr}) and hospital admission id (hadmn_id) and deliver the drugs to the patient.

Algorithm 4.1: Algorithm fixAppointment

Input: S_{id}, D_{addr}, C_C, F_H, P_I, P_{MH}, P_{MS}, P_{al}, P_{icd}
Output: fix appointment

1. Require[D_{addr}][S_{id}]. Patient.length != appointmentslot[doctor][S_{id}].maxpatients
2. Display "Slot is Full"
3. Declare patid as String
4. Set totalPatients = Length of uniquepatientAddress array
5. For I from 0 to totalPatients – 1
6. set Patient = uniquepatientAddress[i]
7. set randpatientId=uniquepatientstring[i]
8. If Patient equals msg.sender
9. set patientid = randpatientid
10. End For
11. Push patientid to appointmentslot[D_{addr}][S_{id}].patient array
12. Create Hospital_Patient_Admission object with hsp_adm with the following properties:
13. hospital_admission_id = hsp_adm
14. _chiefComplaint = C_c
15. _familyhistory = F_H
16. _presentIllness = P_I
17. _pastMedicalHistorg = P_{MH}
18. _majorSurgical = P_{MS}
19. _allergies = P_{al}
20. _icdcodes = P_{icd}
21. prescriptions = Empty array of Prescription objects
22. rand_doctor_assigned_id=doctor
23. slot_id=_slotid
24. purchased_status = false
25. deliver_status = false
26. Push hsp_adm to hospitaladmission[patientid] array
27. End Algorithm

The algorithm 4.1 began by checking the availability of the specified appointment slot for the given doctor, ensuring it was not already filled to its maximum capacity; if it was, an error message was displayed. Subsequently, it searched for the patient ID corresponding to the sender's address among the stored patient addresses, assigning it to patid upon discovery. The algorithm then updated the appointment slot by adding the patient ID to its list of scheduled patients. Following this, it constructed a hospital admission record (hsp_adm) encapsulating various medical details provided as parameters. This record included information such as chief complaint C_C, family history F_H, present illness P_I, past medical history P_{MH}, major surgical history P_{MS}, allergies P_{al}, and an array of ICD codes P_{icd}, along with properties related to prescriptions, Doctor ID, slot ID, and status flags. Finally, the newly created hospital admission record was appended to the existing records under the patient's ID within the hospital admission mapping. Overall, the algorithm mimicked the functionality of the original Solidity function by executing a series of logical steps to facilitate the fixing of appointments and the recording of associated hospital admission data.

Algorithm 4.2: Buy Drugs

Input: Patient, hadmn_id
Output: Buy Drugs

1. Declare admissions as array of Hospital_P_adm objects, retrieved from hospitaladmission mapping using _patient_id
2. Set total_cost to 0
3. For each admission in admissions
4. If admission's rand_hadm_id equals _hadm_id
5. Require admission's purchased_status to be false, else display "Already Purchased the Drugs"
6. For each prescription in admission's prescriptions
7. Retrieve drug ID and quantity from the prescription
8. Increment total_cost by the price of the drug multiplied by its quantity
9. End For
10. Create drugDetailsArray as an empty array of strings with length equal to the number of prescriptions
11. For each prescription in admission's prescriptions
12. Retrieve drug ID
13. Construct drugDetails string containing drug information (ID, name, dose, price, and quantity)
14. Append drugDetails to drugDetailsArray
15. Concatenate all strings in drugDetailsArray into a single string allDrugDetails
16. End For
17. Require msg.value to be greater than or equal to total_cost, else display "Insufficient funds" error
18. Calculate money_to_transfer as total_cost divided by 1 ether
19. Transfer money_to_transfer to the admin address
20. Set admission's purchased_status to true
21. Emit DrugPurchaseReceipt event with the following parameters: _patient_id, _hadm_id, allDrugDetails, total_cost, and false
22. End If
22. End For
23. End Algorithm

The algorithm Buy_Drugs operates by first retrieving the hospital admission records associated with the given patient ID (_patient_id) from the `hospitaladmission` mapping and initializing the `total_cost` variable. It then iterates through each hospital admission record, checking if the provided hospital admission ID (`_hadm_id`) matches the current admission's ID and verifying that the drugs haven't been purchased before. For each prescription within the admission, it calculates the total cost of drugs and constructs a detailed string containing information about each prescribed drug. Subsequently, it checks if the payment sent by the patient is sufficient to cover the total cost, transfers the funds to the admin if so, updates the purchase status, and emits a receipt event with relevant information including patient ID, hospital admission ID, drug details, total cost, and a flag indicating undelivered drugs. This algorithm ensures a systematic process for purchasing drugs within the hospital system, adhering to constraints such as purchase status and payment sufficiency.

Algorithm 4.2: Deliver Drugs

Input: Patient, hadmn_id
Output: Deliver Drugs
1. Declare admissions as array of Hospital_P_adm objects, retrieved from hospitaladmission mapping using _patient_id
2. Set total_cost to 0
3. Require the sender's address (msg.sender) to be present in the uniquechemistAddress array, indicating that only a chemist can mark delivery
4. For each admission in admissions
5. If admission's rand_hadm_id equals _hadm_id
6. Require admission's purchased_status to be true, indicating drugs have been purchased, else display "Drugs Not purchased Yet" error
7. Require admission's deliver_status to be false, indicating drugs have not been delivered, else display "Drugs already delivered" error
8. For each prescription in admission's prescriptions
9. Retrieve drug ID and quantity from the prescription
10. Increment total_cost by the price of the drug multiplied by its quantity
11. End For
12. Create drugDetailsArray as an empty array of strings with length equal to the number of prescriptions
13. For each prescription in admission's prescriptions
14. Retrieve drug ID
15. Construct drugDetails string containing drug information (ID, name, dose, price, and quantity)
16. Append drugDetails to drugDetailsArray
17. Concatenate all strings in drugDetailsArray into a single string allDrugDetails
18. End For
19. Emit DrugPurchaseReceipt event with the following parameters: _patient_id, _hadm_id, allDrugDetails, total_cost, and true (indicating drugs have been delivered)
20. Update admission's deliver_status to true
21. End If
22. End For
23. End Algorithm

The algorithm mark_drugs_delivered begins by retrieving hospital admission records associated with the provided patient ID from the hospitaladmission mapping and initializes the total_cost variable. It verifies that the sender of the transaction is listed as a chemist, ensuring only authorized personnel can mark drug deliveries. It iterates through each hospital admission record, confirming that the provided admission ID matches and that the drugs associated with the admission have been purchased but not yet marked as delivered. For each prescription within the admission, it calculates the total cost of drugs and constructs a detailed string containing information about each prescribed drug. Upon verification of sufficient purchase and undelivered status, it emits a receipt event with relevant details and updates the delivery status of the admission to indicate successful delivery. This algorithm ensures accurate tracking and authorization of drug deliveries within the system.

VI. RESOURCES

Due to the extensive size of the dataset, processing it using a CPU alone was impractical. Therefore, we opt to utilize a GPU, specifically the NVIDIA A4096 GB GPU in conjunction with the ThinkSystem SR560 V2 model equipped with Intel Xeon Silver 4316 40 Core Processors and 256GB RAM, as well as a 4.8 TB HDD. This configuration allows for efficient handling of the datasets computational demands. In the research, **Ganache** software was utilized to support blockchain experimentation, providing a controlled environment for smart contract deployment and testing. Additionally, **TensorFlow** was selected as the primary framework for machine learning investigations. This strategic choice allowed for the exploration of Ethereum-related studies with Ganache and the utilization of TensorFlow robust capabilities for various machine learning tasks within the research framework. Integrating these tools facilitated comprehensive analysis and exploration at the intersection of blockchain and machine learning domains.

VII. EXPERIMENTAL RESULTS & DISCUSSION

Thus, we have created three models and combined these models as a single model to achieve a high accuracy and this also increases the accuracy of the prediction. As we said, we have used three models Convolution LSTM(Conv-LSTM), Bidirectional LSTM(Bi-LSTM) and Segment level Gated Recurrent Unit (Seg-GRU).

The figure 6 despites the architecture of Conv-LSTM. In The Conv-LSTM model the input is first given to the model in the shape of (1 ,14) the given input is then reshaped and provided as the input to the convolution layer which is one dimensional(conv1d) followed by a maxpooling layer for down-sampling or subsampling the spatial dimensions of the input volume which helps to reduce the computational complexity. Then the output

Table 3 Results

	Accuracy	F1	AUPRC (Ȧ)	AUROC (¥)	Hamming Loss(HL)
Conv-LSTM	0.8666	0.86229	0.7771	0.8768	0.0208
Bi-LSTM	0.8666	0.86415	0.7714	0.8777	0.0200
Seg-GRU	0.8666	0.8608	0.7690	0.8774	0.0213
Ensemble Model	0.9333	0.9985	0.9965	0.9981	0.0002

from the maxpooling layer is given to another conv1d layer followed by a dense layer which takes the high-level filtered features form the previous layer and combine it to make prediction. These features were forwarded to a LSTM layer followed by a dropout layer and a dense layer in the shape of(1,306). This novel CNN-LSTM layer provides an accuracy of 0.86 with loss of 0.37.

The figure 7 despites the architecture of Bi-LSTM. The Bi-LSTM model was provided with the input size of (1,14), this input was then reshaped to (1,14,1) and forwarded to a dense layer for taking the high-level filtered features and then to a dropout layer for preventing the model from overfitting and for improving the generalization of the model and it is again provided to dense layer with previous output shape as input and this provides an output shape of (1,14,289). The processed data is given to a Bi-LSTM pretrained model with the input shape of (1,14,289) to get a output shape (1,600) and for getting more general classification the data was given to a dense layer for flattening the input to the shape of (1,19) and By normalizing the layer's input, a batch normalization layer helps to stabilize and quicken the training process. Then the input shape of (1,19) is given to a dense layer and get an output shape(1,306). This model gives an accuracy of 0.86 and lesser loss than Conv-LSTM with 0.02.

The figure 8 despites the architecture of Segment Level GRU. The architecture of segment level gated recurrent unit consists of dense, dropout and batch normalization layers and a flatten layer.The reshaped input of shape (1,14) is given as input to the continuously two dense layers to get a output shape of (1,20), it is then flatten and provided for model and to the batch normalization and at last a dense layer is applied to get the output shape of (1,306). While experimenting the Segment level GRU provided an accuracy of 0.86 with 0.02 loss.

The figure 9 despites the architecture of Ensemble Model. The Ensemble model is the combination of the three models that is Conv-LSTM, Bi-LSTM, Seg-GRU. The Input of shape (1,14) is given to sequential layer of each model which provides output shape of (1,306). The output of each model then concatenates and given as input and get a output shape of (1,918). For reshaping the output from the combined model, the output is then given to a dense layer as input, the dense layer takes the high-level feature from previous layer and combine it for the prediction. The Combined model provides an accuracy of 0.93 with loss 0.0002.

From the table (Table 3) one can infer that by using the combined model the accuracy of the prediction has been increased and hamming loss has also significantly decreased. So, by using these models one can achieve a significant accuracy. Figures 6,7,8 and 9 describe the architecture of the respective models.

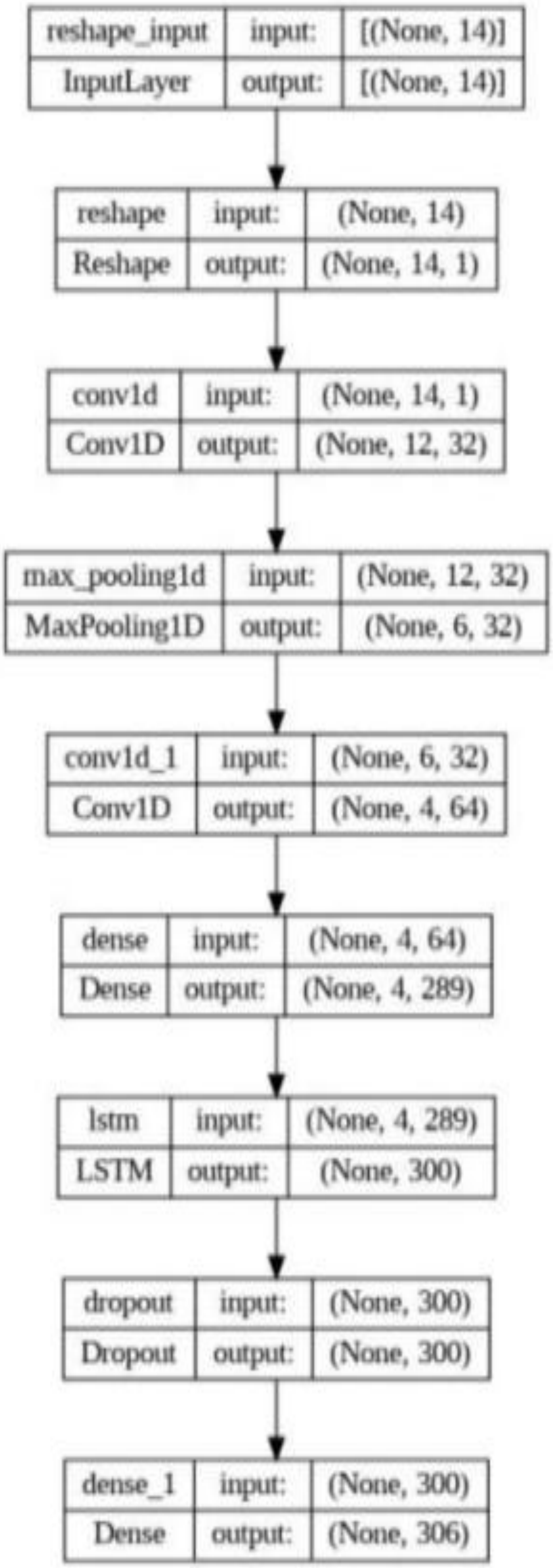


Fig. 6 Conv-LSTM Architecture

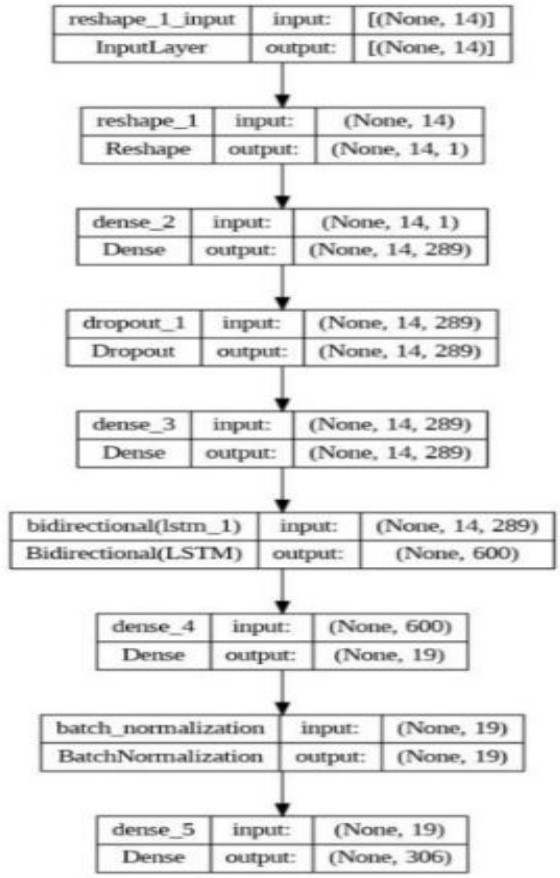


Fig. 7 Bi-LSTM Architecture

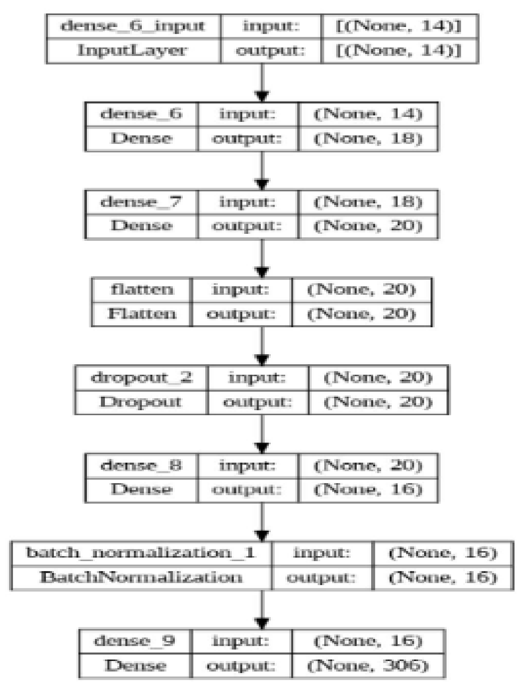


Fig. 8 Seg-GRU Architecture

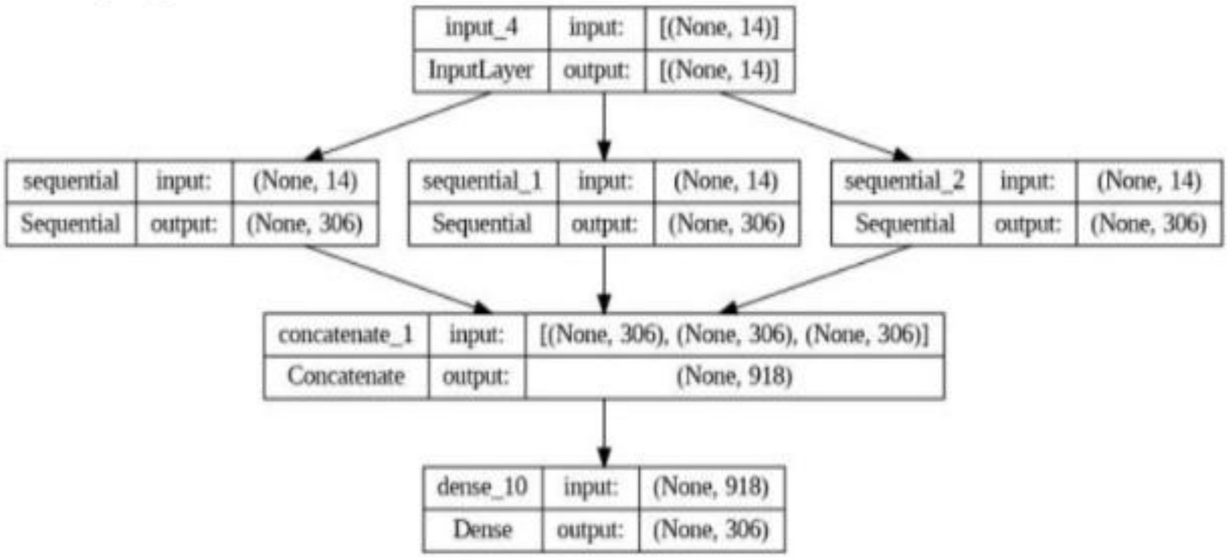


Fig. 9 Ensemble Model Architecture

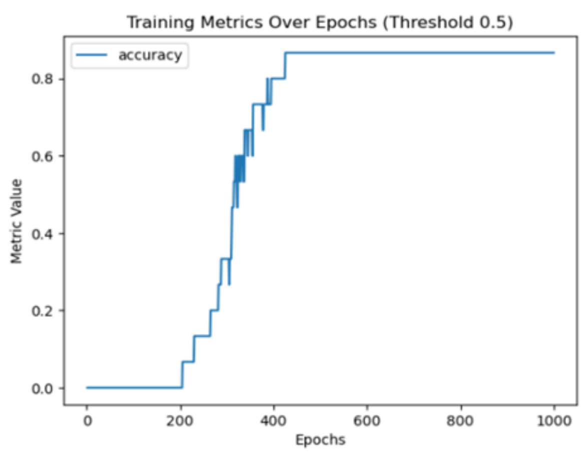


Fig.10 Conv-LSTM Accuracy Graph

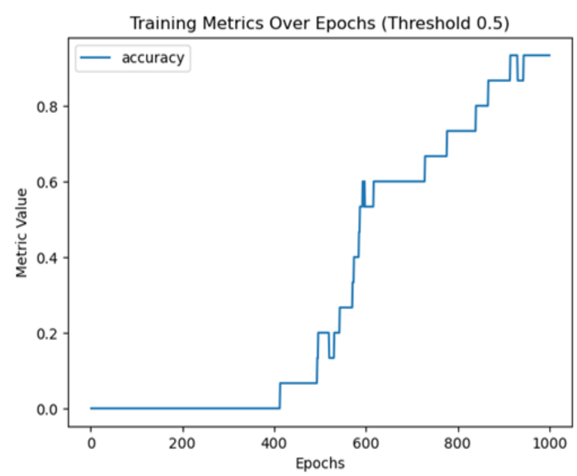
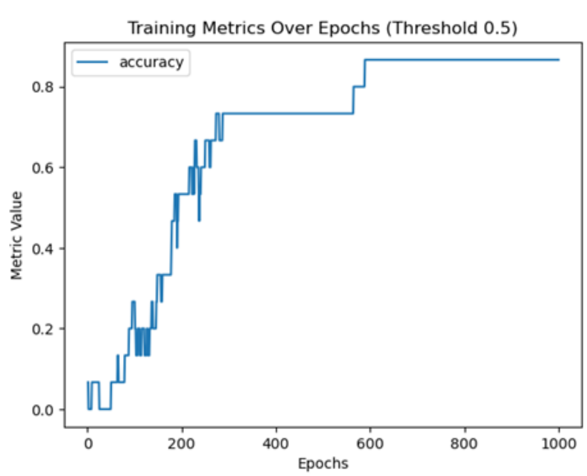
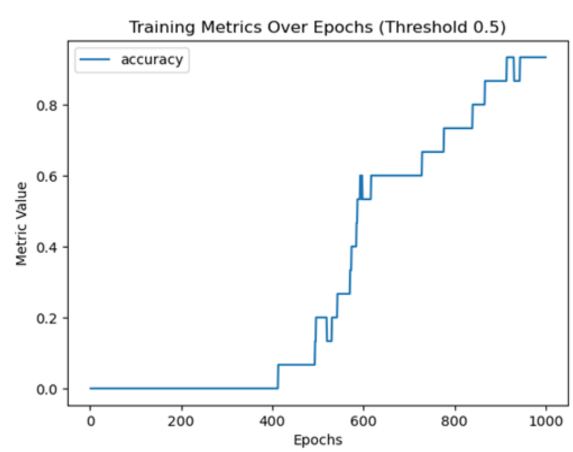


Fig.11 Bi-LSTM Accuracy Graph





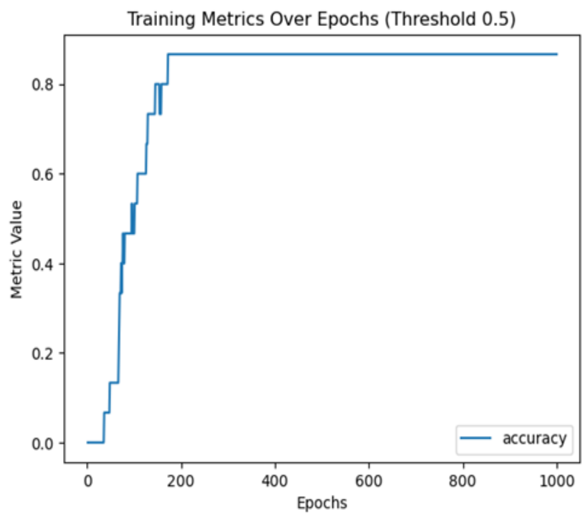


Fig.12 Seg-GRU Accuracy Graph

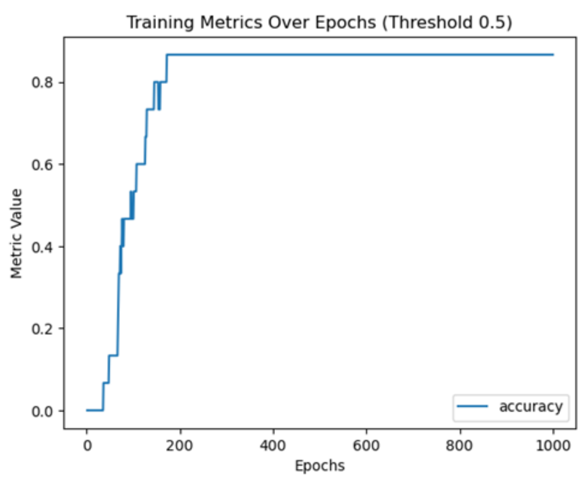


Fig.13 Ensemble Model Accuracy Graph

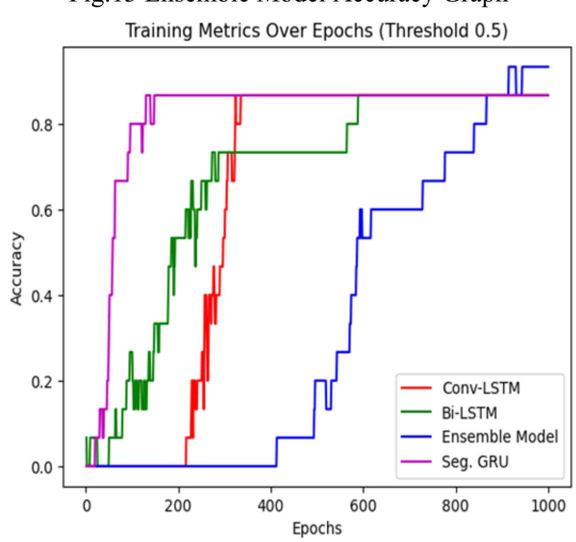


Fig.14 Combined Accuracy Graph

The graphs depicted in Fig. 10, Fig. 11, and Fig. 12 demostrate the accuracy achieved at each epoch, with a threshold set at 0.5.The accuracy attained by the conv-LSTM model, as depicted in Figure 9, exhibits a continuous increase from epoch 200 onwards. By approximately epoch 400, it surpasses 86.66% and remains stable throughout the remaining epochs. Figure 10 depicts the accuracy achieved by the Bi-LSTM model. Unlike the Conv-LSTM model, Bi-LSTM demonstrates an increase in accuracy starting from epoch 300, with fluctuations in subsequent epochs. It surpasses 86% accuracy at epoch 800 and maintains this level throughout the remaining epochs. The Seg-GRU graph from Figure 11 shows a gradual increase in accuracy starting from epoch 350, ultimately reaching 86.66%. In contrast, the ensemble model accuracy, as illustrated in the accompanying figure, exhibits a rapid increase from the onset, reaching nearly 93% before epoch 200, and remaining stable for the subsequent epochs.

Table 4 Block 569 Details

Parameters	Values
Total Transactions	1
Size of Block	552 bytes
Gas Used	89301
Gas Limit	6721975
Time Stamp	2024-03-14 19:57:16
Block Hash	0x42a625f43ecfdbe81f508abbb5cf36455c8ae544d45c9222cd9c7e59aaba1aff

Table 5 Block 570 details

Parameters	Values
Total Transactions	1
Size of Block	552 bytes
Gas	107751

Used	
Gas Limit	6721975
Time Stamp	2024-03-14 19:57:24
Block Hash	0x87abcd94f526506af4bd3837e88493961cc12f89a90692cf23f636daad2ec0c4

The analysis of Blocks 569 and 570 from the blockchain reveals several noteworthy observations. Both blocks contain a single transaction and maintain a consistent block size of 552 bytes, reflecting efficiency and uniformity in the mining process. Despite the identical timestamp of March 14, 2024, at 19:57:16, and March 14, 2024, at 19:57:24, respectively, there is a discernible increase in gas usage from Block 569 to Block 570, with gas values of 89,301 and 107,751, respectively. This uptick suggests a potential escalation in computational complexity or transaction intricacy within the blockchain network. Notably, the gas limit remains unchanged at 6,721,975, indicating stability in network parameters or policies governing gas limits across these blocks. The inclusion of only one transaction per block may imply either a relatively low volume of transactions during this period or a deliberate strategy to prioritize certain transactions for inclusion in blocks.Furthermore, the unique block hashes for Block 569 and Block 570, underscore the cryptographic integrity and individuality of each block within the blockchain. These insights provide valuable perspectives on the network's performance, scalability, and transaction dynamics during the analyzed time period.

VIII. CONCLUSION

The research not only addresses the pressing need for digitizing EHRs but also tackles the inherent challenges associated with traditional storage methods. By integrating blockchain technology, the framework ensures data integrity, security and accessibility while decentralizing healthcare management. Furthermore, the deep learning models employed for feature extraction and prediction demonstrate promising accuracy rates of 93% in disease prediction, paving the way for personalized and evidence-based healthcare solutions. The framework emphasis on patient-centric care, interoperability and transparency signifies a significant step towards revolutionizing healthcare systems globally. Moreover, the proposed framework not only streamlines processes such as patient enrollment, appointment scheduling and prescription management but also facilitates seamless collaboration among healthcare stakeholders. The utilization of Ethereum blockchain for secure communication and data storage ensures patient privacy and confidentiality addressing critical concerns in the healthcare domain. By harnessing the power of blockchain and deep learning technologies healthcare organizations can extract new information from enormous volumes of unstructured EHR data, leading to more informed decision-making and improved patient care pathways. Looking ahead, further research and development efforts were warranted to refine and optimize the framework for broader adoption in real-world healthcare settings. In conclusion, the novel cognitive blockchain-based framework represents a significant step towards revolutionizing healthcare management by leveraging cutting-edge technologies to address longstanding challenges. By fostering secure, transparent, and patient-centric healthcare ecosystems, the framework has the potential to drive positive outcomes for both healthcare providers and patients alike. Continued collaboration, innovation, and adoption of such transformative frameworks are essential for ushering in a new era of data-driven, personalized healthcare delivery. Additionally, ongoing advancements in blockchain technology, deep learning algorithms and data privacy protocols offer opportunities for enhancing the framework capabilities and expanding its scope of application.

Conflict of interest: The authors declare that there are no conflicts of interest.
Competing Interest:
The authors declare that they have no competing interests

Author Contribution:
All authors contributed equally to the design, data collection, analysis, implementation and evaluation of models and manuscript preparation.

Data Availability Statement
The data used in this research can be downloaded from https://physionet.org/content/mimiciv/3.1/ after completing the modules in the website.

Funding Information
The authors declare that no funding was received for this research.

Research Involving Human and /or Animals
Not Applicable

Informed Consent Not Applicable

REFERENCES

[1] Tushaar Gangavarapu, Gokul S Krishnan, Sowmya Kamath S And Jayakumar Jeganathan,"FarSight: Long-Term Disease Prediction Using Unstructured Clinical Nursing Notes,"IEEE Transactions On Emerging Topics In Computing, Volume 9, No. 3, July-Sept. 2021.
[2] Xiaowen Zhang, Yuxia Lei, Guangshun Li, Shi Wang, Huifeng Li,"Combining Reverse Temporal Attention Mechanism and Dynamic Similarity Analysis for Disease Prediction",IEEE International Conference on Dependable, Autonomic and Secure Computing,25 Dec 2023.
[3] Shang Gao, Mohammed Alawad, M. Todd Young, John Gounley, "Limitations of Transformers on Clinical Text Classification",IEEE Journal of Biomedical and Health Informatics,Volume 25, No. 9, September 2021.
[4] Ning Liu, Qian Hu, Huayun Xu, Xing Xu, Mengxin Chen,"Med-BERT: A Pretraining Framework for Medical Records Named Entity Recognition", IEEE Transactions on Industrial Informatics, Volume 18, No. 8, August 2022
[5] Cherubin Mugisha, Incheon Paik,"Comparison of Neural Language Modeling Pipelines for Outcome Prediction From Unstructured Medical Text Notes",IEEE Access,Volume. 10,31 Jan 2022
[6] Minji Kim, Inwhee Joe, "Automatic Diagnosis of Medical Conditions Using Deep Learning With Symptom2Vec", IEEE Access,Volume. 11,13 Oct 2023
[7] Shunli Zhang, Yancui Li, Shiyong Li, Fang Yan, "Bi-LSTM-CRF Network for Clinical Event Extraction With Medical Knowledge Features",IEEE Access,Volume. 10,10 Oct 2022
[8] Xiao Luo, Priyanka Gandhi, Susan Storey, Kun Huang,"A Deep Language Model for Symptom Extraction From Clinical Text and its Application to Extract COVID-19 Symptoms From Social Media", IEEE Journal of Biomedical and Health Informatics,Volume. 26, No. 4, April 2022
[9] Yash Pawar, Aron Henriksson, Pontus Hedberg, Pontus Naucler,"Leveraging Clinical BERT in Multimodal Mortality Prediction Models for COVID-19", 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS),31 Aug 2022
[10] Mingming Lu, Yu Fang, Fengqi Yan, Maozhen Li,"Incorporating Domain Knowledge into Natural Language Inference on Clinical Texts",IEEE Access, Volume: 7,29 April 2019
[11] Syed Atif Moqurrab, Umair Ayub, Adeel Anjum, Sohail Asghar, Gautam Srivastava,"An Accurate Deep Learning Model for Clinical Entity Recognition From Clinical Notes", IEEE Journal of Biomedical and Health Informatics,Volume. 25, No. 10, October 2021
[12] Dengao Li, Huiting Ma, Wenjing Li, Baofeng Zhao, Jumin Zhao, Yi Liu, Jian Fu,"KTI-RNN: Recognition of Heart Failure from Clinical Notes",Tsinghua Science and Technology, Volume. 28, No. 1, February 2023
[13] Ying Yu, Min Li,Liangliang Liu, Yaohang Li, Jianxin Wang,"Clinical big data and deep learning: Applications, challenges, and future outlooks", Big Data Mining and Analytics ,Volume, 2,No. 4, Dec 2019
[14] Gaspard Harerimana, Jong Wook Kim, Hoon Yoo, Beakcheol Jang,"Deep Learning for Electronic Health Records Analytics", IEEE Access,Volume: 7,11 July 2019
[15] Fei Teng, Yiming Liu, Tianrui Li, Yi Zhang, Shuangqing Li, Yue Zhao,"A Review on Deep Neural Networks for ICD Coding",IEEE Transactions on Knowledge and Data Engineering,Volume. 35, No. 5, 01 May 2023
[16] Saddam Al Amin, Md. Saddam Hossain Mukta, Md. Sezan Mahmud Saikat, Md. Ismail Hossain, Md. Adnanul Islam,"Data Driven Classification

of Opioid Patients Using Machine Learning–An Investigation",IEEE Access ,Volume. 11,19 Dec 2022

[17] Md Nahid Hasan, Sammi Hamdan, Samir Poudel, Jorge Vargas, Khem Poudel,"Prediction of Length-of-stay at Intensive Care Unit (ICU) Using Machine Learning based on MIMIC-III Database",2023 IEEE Conference on Artificial Intelligence (CAI),02 Aug 2023

[18] Yirui Dong,"A Deep Learning-Driven Disease Classification Method Using MIMIC-III Database and Natural Language Processing", 2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA), 23 Jan 2024

[19] Yongsik Jin, Crino Shin, Wookyong Kwon, Kyuhyung Kim, Jong Pil Yun,"Data filtering for corrupted MIMIC III dataset with deep learning",2020 20th International Conference on Control, Automation and Systems (ICCAS),01 Dec 2020

[20] Dillon Chrimes, Chanhee Kim,"Comparison of MIMIC-III and MIMIC-IV for big data analytics of health informatics",2023 IEEE International Conference on Big Data (BigData),22 Jan 2024

[21] Serdar Arslan,"Application of BiLSTM-CRF model with different embeddings for product name extraction in unstructured Turkish text",Journal of Neural Computing and Applications,Feb 2024.

[22] Joy Gorai,Dilip Kumar Shaw,"A BERT-encoded ensembled CNN model for suicide risk identification in social media posts",Journal of Neural Computing and Applications,Mar 2024

[23] Akhilendra Pratap Singh, Nihar Ranjan Pradhan, Ashish K. Luhach, Sivansu Agnihotri, Noor Zaman Jhanjhi, Sahil Verma,"A Novel Patient-Centric Architectural Framework for Blockchain-Enabled Healthcare Applications",IEEE Transactions on Industrial Informatics, Volume. 17, No. 8, Aug 2021

[24] Sabri Barbaria, Marco Casassa Mont, Essam Ghadafi, Halima Mahjoubi Machraoui, Hanene Boussi Rahmouni,Leveraging Patient Information Sharing Using Blockchain-Based Distributed Networks,IEEE Access,Volume. 10, 12 Sep 2022

[25] Kebira Azbeg, Ouail Ouchetto, Said Jai Andaloussi,"Access Control and Privacy-Preserving Blockchain-Based System for Diseases Management",IEEE Transactions on Computational Social Systems,Volume. 10, No. 4, August 2023.

[26] Guangyu Peng, Aiqing Zhang, Xiaodong Lin,"Patient-Centric Fine-Grained Access Control for Electronic Medical Record Sharing With Security via Dual-Blockchain",IEEE Transactions on Network Science and Engineering, Volume: 10, No: 6, Nov.-Dec. 2023

[27] Jingwei Liu, Yue Fan, Rong Sun, Lei Liu,"Blockchain-Aided Privacy-Preserving Medical Data Sharing Scheme for E-Healthcare System", IEEE Internet of Things Journal, Volume. 10, No. 24, 15 December 2023

[28] Sadia Ramzan, Aqsa Aqdus, Vinayakumar Ravi, Deepika Koundal, Rashid Amin, Mohammed A. Al Ghamdi,"Healthcare Applications Using Blockchain Technology: Motivations and Challenges",IEEE Transactions on Engineering Management,Volume: 70, Issue: 8, August 2023

[29] Yu Jiang, Xiaolong Xu, Fu Xiao,"Attribute-Based Encryption With Blockchain Protection Scheme for Electronic Health Records",IEEE Transactions on Network and Service Management,Volume: 19, Issue: 4, December 2022

[30] Aysha Alnuaimi, Amna Alshehhi, Khaled Salah, Raja Jayaraman, Ilhaam A. Omar, Ammar Battah,"Blockchain-Based Processing of Health Insurance Claims for Prescription Drugs", IEEE Access ,Volume: 10,04 November 2022

[31] Kun Zhou, Chunping Wang, Xiaoqing Wang, Shenghang Chen, Hongbing Cheng,"A Novel Scheme to Improve the Scalability of Bitcoin Combining IPFS With Block Compression", IEEE Transactions on Network and Service Management,Volume: 19, Issue: 4, December 2022

[32] Wei Liang, Yang Yang, Ce Yang, Yonghua Hu, Songyou Xie, Kuan-Ching Li, Jiannong Cao,"PDPChain: A Consortium Blockchain-Based Privacy Protection Scheme for Personal Data", IEEE Transactions on Reliability,Volume: 72, Issue: 2, June 2023

[33] Song Tang, Zhiqiang Wang, Jia Dong, Yandong Ma,"Blockchain-Enabled Social Security

Services Using Smart Contracts",IEEE Access,Volume: 10,14 July 2022

[34] Sumit Kumar Rana, Arun Kumar Rana, Sanjeev Kumar Rana, Vishnu Sharma, Umesh Kumar Lilhore, Osamah Ibrahim Khalaf,"Decentralized Model to Protect Digital Evidence via Smart Contracts Using Layer 2 Polygon Blockchain", IEEE Access (Volume: 11),07 August 2023

[35] Liwei Ouyang, Fei-Yue Wang, Yonglin Tian, Xiaofeng Jia, Hongwei Qi, Ge Wang,"Artificial Identification: A Novel Privacy Framework for Federated Learning Based on Blockchain",IEEE Transactions on Computational Social Systems,Volume: 10, Issue: 6, December 2023

[36] Kai Zhang, Yan Zhang, Yanping Li, Ximeng Liu, Laifeng Lu,"A Blockchain-Based Anonymous Attribute-Based Searchable Encryption Scheme for Data Sharing", IEEE Internet of Things Journal ,Volume: 11, Issue: 1, 01 January 2024

[37] Mikail Mohammed Salim, Jong Hyuk Park,"Federated Learning-Based Secure Electronic Health Record Sharing Scheme in Medical Informatics", IEEE Journal of Biomedical and Health Informatics,Volume: 27, Issue: 2, February 2023

[38] Saminur Islam, Mohammad Jaminur Islam, Mahmud Hossain, Shahid Noor, Kyung-Sup Kwak, S. M. Riazul Islam,"A Survey on Consensus Algorithms in Blockchain-Based Applications: Architecture, Taxonomy, and Operational Issues",IEEE Access,Volume: 11,13 April 2023

[39] Murat Kuzlu, Manisa Pipattanasomporn, Levent Gurses, Saifur Rahman,"Performance Analysis of a Hyperledger Fabric Blockchain Framework: Throughput, Latency and Scalability",2019 IEEE International Conference on Blockchain (Blockchain),02 January 2020

[40] Ivan Laishevskiy, Artem Barger, Vladimir Gorgadze,"A Journey Towards the Most Efficient State Database for Hyperledger Fabric", 2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 12 July 2023

[41] Munish Katoch, Kavita Devi, Anuj Sharma, Prikshat Kumar Angra, Pritpal Singh, Ashwani Kumar,"Enhancing the Security of E-Healthcare Transactions Using Hyperledger Fabric Modular Blockchain Framework",2023 6th International Conference on Contemporary Computing and Informatics (IC3I),26 January 2024

[42] N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson,and T. Hayajneh, “Healthcare blockchain system using smart contracts for secure automated remote patient monitoring,” J. Med. Syst., vol. 42, no. 7, 2018