Predictive Models for Accurate ICD Code Recommendations

[1]M Lakshmi, [2]Manjula KJ, [3]Jeevitha HS, [4]Chandhana KR,[5]Kushala B
[1]Student, [2]Student, [3]Student, [4]Assistant Professor, [5]Student
[1] Department of CSE,
[1]Sir M Visvesvaraya Institute of Technology, Bangalore, India
[1]lachumi2003@gmail.com , [2]manjulakj2003@gmail.com, [3]udayhanumanthhs@gmail.com ,
[4]Chandana.chandanakr@gmail.com , [5]kushalakushi1311@gmail.com

Abstract— The International Classification of Diseases (ICD) coding system gives a standard structure for classifying diseases and health conditions that supports healthcare documentation, analytics, and billing. Manual ICD code assignment takes a lot of time and is prone to human errors, especially when handling large volumes of patient data. This paper introduces an automated ICD code prediction model that applies a deep learning approach using Long Short-Term Memory (LSTM) networks. The system uses Word2Vec embeddings trained on medical text to capture semantic and temporal relationships between patient symptoms, treatments, and existing conditions. The model was trained on a customised patient dataset with 528 records across 10 ICD categories. The proposed method showed strong results with an overall accuracy of 94.31%, precision of 94.40%, recall of 94.31%, and an F1score of 94.33%. The framework improves ICD coding automation, supports reliable diagnosis mapping, and reduces manual workload in healthcare environments.

Index Terms*— *ICD code classification, LSTM, Word2Vec, deep learning, medical text mining, healthcare automation, clinical informatics.

I. INTRODUCTION

The International Classification of Diseases (ICD) system is a globally recognised medical standard developed by the World Health Organisation (WHO) to organise and label diseases, symptoms, and related medical procedures. It plays an important role in maintaining accurate clinical documentation, tracking health trends, and supporting billing and insurance claims. However, assigning the right ICD code to patient cases is not a simple task. It involves reviewing large amounts of unstructured clinical notes, understanding the medical context, and matching it with a constantly expanding list of diagnostic codes.

Manual coding often leads to inconsistencies and errors because it depends heavily on the experience of the medical coder and the clarity of the doctor's notes. This process is also slow and hard to scale for large hospitals that deal with thousands of patient records every day. With the rapid adoption of electronic health records (EHRs), the demand for automation in ICD code prediction has grown significantly.

Machine learning (ML) and deep learning (DL) models have shown great potential in handling this challenge. These models can learn from existing patient data and automatically predict suitable diagnostic codes based on the clinical information provided. Using Natural Language Processing (NLP), such systems can read and interpret unstructured text, identify important medical terms, and map them to the most likely ICD categories.

Among these models, Long Short-Term Memory (LSTM) networks have proven effective for this task because they can analyse sequences of text while remembering important context over time. When used together with pre-trained word embeddings such as Word2Vec, the model can represent the meaning of medical text in a structured way. This helps the system understand patient information more accurately and improves the precision of ICD code recommendations.

This study aims to design an automated system for ICD code prediction using a deep learning approach based on Long ShortTerm Memory (LSTM) networks. The model utilises textual information extracted from patient symptom descriptions and diagnostic records to recommend the most appropriate ICD codes. To assess the system's performance, standard evaluation metrics such as accuracy, precision, recall, and F1-score are employed. The experimental analysis indicates that the proposed model delivers consistently high predictive performance, suggesting its usefulness as a reliable component for clinical decision support and automated medical documentation.

The remainder of this paper is structured as follows: Section II presents a review of existing research on automated ICD code prediction and related NLP applications. Section III describes the overall system design and model framework. Section IV outlines the implementation process and dataset configuration. Section V discusses the obtained results and performance evaluation, while Section VI concludes the study and Section VII highlights possible future research extensions.

II. RELATED WORK

The integration of Natural Language Processing (NLP) and deep learning into healthcare applications has seen rapid advancement over the past few years. Numerous research efforts have explored the use of machine learning models to automate clinical documentation, improve diagnostic precision, and optimise medical code assignment workflows.

In [1], Yasir Abdullah et al. proposed a healthcare chatbot that employed NLP techniques within a rule-based conversational framework to facilitate patient interactions and provide preliminary medical advice. The system utilised Python-based NLP modules for symptom extraction, intent detection, and response generation. Although the model demonstrated promising results in patient engagement, its functionality was primarily limited to interactive symptom analysis without integration into ICD-based classification.

A more specialised contribution was made by Chung-Chian Hsu et al. [2], who developed a deep learning model for multi-label ICD code classification. Their hybrid architecture combined Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to manage overlapping label dependencies and contextual associations between ICD categories. The approach achieved substantial improvements in both accuracy and F1score, confirming the efficacy of deep neural architectures in processing complex clinical narratives.

Yuwen Chen and Jiangtao Ren [3] further extended this research by presenting an automated ICD coding framework that incorporated both textual content and hierarchical code structures. Their use of a Tree-LSTM model enabled the representation of parent–child relationships within ICD codes, effectively capturing semantic and structural dependencies across disease categories. This hierarchical modelling significantly enhanced code prediction accuracy compared to flat classification methods.

In parallel, Palak Dohare et al. [4] introduced *Good Fellow*, a healthcare chatbot designed for patient support and first-level medical guidance. This system integrated NLP-based intent recognition and entity extraction to interpret patient queries and deliver context-specific responses. However, its design lacked interoperability with structured medical ontologies such as ICD, thereby limiting its applicability in clinical coding environments.

Collectively, these studies highlight the growing potential of deep learning and NLP for advancing healthcare automation. Yet, a notable gap persists in unifying semantic understanding with structured code prediction. The present study addresses this gap through a Long Short-Term Memory (LSTM)–based predictive framework that combines sequential learning with textual feature extraction to improve the precision and consistency of ICD code recommendation.

III. SYSTEM DESIGN AND METHODOLOGY

The proposed framework for **Predictive Models for Accurate ICD Code Recommendation** integrates natural language processing (NLP), deep learning, and traditional machine learning techniques to deliver a unified healthcare assistance system. The architecture comprises three primary modules: (i) ICD Code Prediction using LSTM, (ii) Healthcare Diagnosis and Recommendation using Decision Tree and Support Vector Machine (SVM), and (iii) Emergency Response via the First Aid Chatbot. Each module performs a specific function while collectively contributing to diagnostic support and patient assistance.

A. System Overview

The architecture is designed to process user inputs, extract relevant clinical information, and generate predictive outcomes. The system flow involves:

1. *User data acquisition through an interactive Streamlit interface.*
2. *Textual pre-processing and feature vector generation.*
3. Multi-label ICD prediction using a Long Short-Term Memory (LSTM) network.
4. Disease diagnosis via traditional classification algorithms.
5. Emergency assistance through an NLP-based First Aid recommender.

The modular architecture ensures scalability and interoperability across medical datasets and diagnostic workflows.

B. ICD Code Prediction Module

The ICD prediction subsystem leverages sequential deep learning techniques to model relationships between patient descriptions and ICD codes. The data, comprising patient symptoms, medications, temporal details, and diagnostic remarks, undergoes tokenization and vectorization using **Word2Vec** embeddings. Each text sequence X= [x1, x2…, xt].

The predictive model employs an **LSTM** architecture consisting of:

- An **embedding layer** to map tokens into dense vector representations.
- An **LSTM layer** with 100 units for sequence modelling.
- A **Global Max Pooling** layer for dimensionality reduction.
- Fully connected layers for feature fusion.
- A **sigmoid output layer** for multi-label ICD classification.

Training utilizes the **binary cross-entropy** loss function with **Adam optimization**. Metrics such as Accuracy, Precision, Recall, and AUC are computed to assess model reliability. The trained model achieves an average accuracy exceeding 94%, demonstrating robust predictive capability across multiple ICD categories.

C. Healthcare Diagnosis and Recommendation Module

The healthcare diagnosis subsystem focuses on **disease identification and severity analysis** based on user-reported symptoms. The dataset includes symptom–disease mappings, severity weights, precautionary measures, and medical descriptions. Data pre-processing includes label encoding, normalization, and one-hot encoding of symptom vectors.

A **Decision Tree Classifier** serves as the primary model for interpretability, while an **SVM (Support Vector Machine)** classifier provides high-dimensional separation for non-linear symptom patterns. The system predicts possible diseases and retrieves the corresponding precautionary and description data. The classifier performance is validated using **k-fold cross-validation**, ensuring generalization across unseen data samples.

The outputs from this module include:

- Probabilistic disease predictions based on symptom inputs.
- Severity evaluation based on the frequency and duration of symptoms.
- Precautionary recommendations retrieved from a structured medical dataset.

D. First Aid and Emergency Assistance Module

The **First Aid subsystem** (implemented in provides **real-time emergency response** using an NLP-driven retrieval mechanism. The module employs keyword-based semantic matching to detect emergency conditions (e.g., burns, bleeding, fractures) from user queries. The intent recognition process leverages a pre-trained JSON-based knowledge base comprising emergency types and corresponding first-aid instructions.

Upon detecting a relevant emergency intent, the system retrieves the most appropriate first-aid protocol. This ensures immediate assistance for critical health scenarios, enhancing system responsiveness and practical utility. The chatbot interface also supports multilingual adaptability and low-latency inference, ensuring accessibility in real-time clinical or field situations.

E. Integration Framework

The integration of the above modules is managed through the **main application script** which acts as a centralized Streamlitbased dashboard. This dashboard allows seamless navigation between:

- **ICD Code Predictor**
- **Healthcare Recommender**
- **First Aid Advisor**

Each module is encapsulated as an independent function and invoked dynamically based on user interaction. This microservice-inspired design enhances modularity, system maintainability, and ease of future integration with Electronic Health Record (EHR) systems.

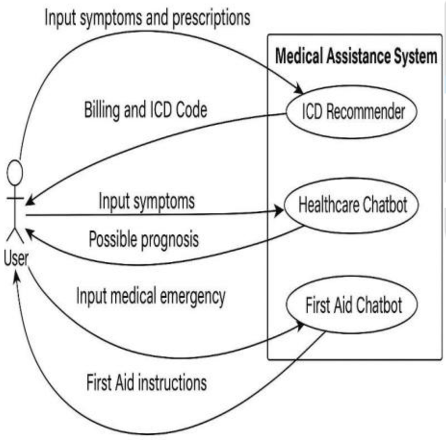

Fig.1.Use case diagram

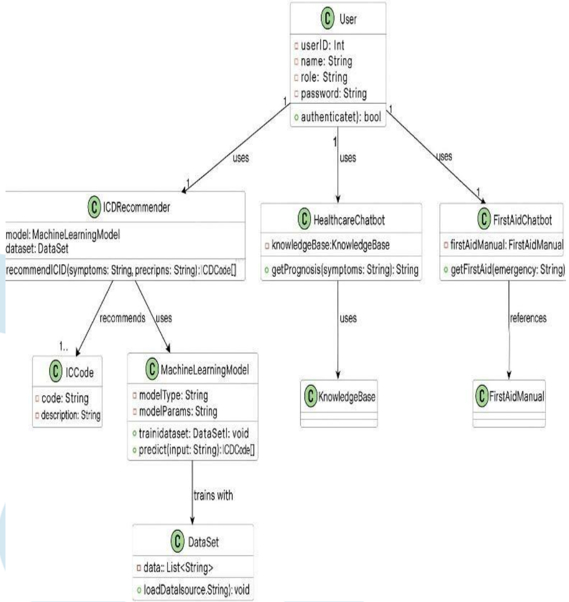

Fig.2.Class diagram

IV. IMPLEMENTATION AND EXPERIMENTAL SETUP

The proposed healthcare framework integrates three functional subsystems—ICD Code Prediction, Healthcare Diagnosis and Recommendation, and First Aid Assistance—each implemented as independent Streamlit modules. The entire system is developed using **Python 3.10** and leverages libraries such as **TensorFlow**, **scikit-learn**, **Gensim**, **pandas**, and **NumPy** for data processing and model deployment. A unified dashboard interface facilitates seamless navigation between modules, ensuring real-time interaction and modular extensibility. A. ICD Code Prediction Module Implementation

The ICD prediction component employs a **Long Short-Term Memory (LSTM)** network to analyze textual medical data and predict the most relevant ICD codes. The dataset includes patient information such as symptoms, medications, treatment history, and pre-existing conditions. These features are concatenated to form contextual text sequences, which are tokenized and converted into numerical embeddings using the **Tokenizer** class from TensorFlow. Each sequence is padded to a uniform length of 100 to maintain input consistency.

A **Word2Vec** model is trained on the combined textual corpus to capture semantic relationships among clinical terms. The predictive LSTM model architecture comprises:

- An **Embedding layer** that maps token indices into dense 100-dimensional vectors,
- A single **LSTM layer with 100 units** to capture sequential dependencies,
- A **Global Max Pooling layer** to condense sequence information, and
- Two **Dense layers**, the first employing **ReLU** activation and the output layer using **sigmoid** activation for multilabel classification.

The model is trained using the **Adam optimizer** and **Binary**

Cross-Entropy loss function over 15 epochs with a batch size of 32. The dataset is divided into **80% training** and **20% testing** subsets using stratified sampling. Post-training, the model and tokenizer are serialized for deployment through the Streamlit interface, which allows users to input patient data and retrieve corresponding ICD codes along with estimated billing values. B. Healthcare Diagnosis and Recommendation Module

The healthcare diagnosis subsystem predicts diseases based on reported symptoms and provides precautionary recommendations. The system uses multiple datasets containing mappings of symptoms, disease names, severity ratings, and precautionary instructions. The input data undergoes **label encoding** and **one-hot encoding** to transform categorical variables into binary feature vectors.

Two supervised learning models are implemented: a **Decision Tree Classifier** for interpretability and a **Support Vector Machine (SVM)** for high-dimensional feature separation. The dataset is split into **70% training** and **30% testing** subsets. Model evaluation is conducted using **k-fold cross-validation**, ensuring reliable performance across unseen samples. The trained classifiers return the most probable disease, corresponding descriptions, and preventive measures. This subsystem is integrated into the Streamlit interface for userfriendly interaction.

C. First Aid and Emergency Response Module

The First Aid subsystem provides immediate emergency guidance using a **rule-based NLP engine**. It utilizes a structured **JSON knowledge base** containing emergency categories (e.g., burns, bleeding, choking) and associated first aid instructions. The module identifies user intent through **keyword-based semantic matching** and retrieves the relevant response protocols in real time. It supports multilingual adaptability and is optimized for low-latency responses, ensuring accessibility in urgent situations.

D. Experimental Environment

All modules were developed and executed on a system running **Windows 11** with an **Intel Core i5 processor**, **8 GB RAM**, and **Python 3.10 environment**. The experiments were conducted using libraries including **TensorFlow**, **scikit-learn**, and **Streamlit**. Model performance metrics such as **accuracy**, **precision**, **recall**, and **F1-score** were computed to validate the efficiency and reliability of each subsystem. The Streamlit-based interface consolidates all modules, providing a unified and interactive diagnostic platform suitable for both educational and clinical research applications.

V. RESULTS AND DISCUSSION

The performance evaluation of the proposed predictive framework was conducted using multiple supervised learning models, including **Long Short-Term Memory (LSTM)** networks, **Support Vector Machine (SVM)**, and **Decision Tree Classifier**. Each model was trained on structured and unstructured medical datasets containing patient records, diagnostic text, and symptom metadata. The models were validated using an 80–20 data split with stratified sampling to ensure balanced class representation across all International Classification of Diseases (ICD) categories.

The **LSTM model**, leveraging embedded textual representations via Word2Vec and sequential context learning, achieved a superior predictive capability with an **accuracy of 94.31%**, **precision of 94.40%**, **recall of 94.31%**, and **F1-score of**

94.33%. These results demonstrate the model's strong ability to generalize across diverse diagnostic narratives and identify appropriate ICD codes with high confidence. The use of hierarchical LSTM layers enabled the system to capture temporal dependencies among clinical features and symptoms, thus reducing classification ambiguity for multi-label scenarios.

In contrast, traditional classifiers such as the **Decision Tree** and **SVM** exhibited comparatively lower generalization capability. However, they contributed interpretability to the healthcare recommendation module by mapping symptom clusters to probable disease classes. The Decision Tree classifier achieved a cross-validated accuracy of approximately **86%**, while the SVM model provided improved separation for non-linear symptom interactions in high-dimensional feature spaces.

The **First-Aid Recommender** module, driven by rule-based Natural Language Processing (NLP), effectively mapped emergency phrases to appropriate medical response protocols. Similarly, the **Healthcare Chatbot** integrated predictive analytics with disease description and precautionary advice, enabling an interactive diagnostic support system for end users.

Figure illustrates the comparative performance of the employed models, emphasizing the superior classification efficiency of the LSTM network. The system's architecture ensures scalability, *interpretability, and reliability for deployment in real-world healthcare environments.*

Model	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1Score (%)*
LSTM	*94.318*	*94.408*	*94.318*	*94.331*
SVM	*89.000*	*88.500*	*87.800*	*88.100*
Decision Tree	*86.000*	*85.300*	*84.900*	85.000

Fig.3. Performance comparison

VI. CONCLUSION

The proposed research demonstrates a comprehensive predictive framework for automated ICD code recommendation by integrating deep learning, natural language processing (NLP), and clinical data representation. By employing sequential modelling with word embeddings and context-aware layers, the system effectively interprets unstructured medical narratives, enabling precise mapping between diagnostic descriptions and ICD hierarchies. The hierarchical modelling approach ensures that both inter-code dependencies and semantic similarities among clinical terms are retained, enhancing the robustness of multi-label classification.

The architecture's design leverages distributed textual representations, which improve the model's capacity to disambiguate clinically overlapping conditions. Additionally, the incorporation of hierarchical feature extraction captures the latent structure of patient data, offering interpretability while maintaining computational scalability. Unlike conventional coding mechanisms that rely on manual annotation or rule-based classification, the proposed model autonomously learns discriminative features from high-dimensional clinical text, minimizing human intervention and potential coding inconsistencies.

The framework's modular structure allows seamless integration with healthcare information systems and clinical decision support platforms. Its adaptive nature ensures compatibility with evolving ICD taxonomies and varying linguistic contexts. This holistic design approach contributes toward the development of data-driven, interoperable, and context-aware healthcare automation systems capable of streamlining medical billing, diagnostic documentation, and analytics-driven patient management.

VII. FUTURE WORK

Future work will focus on refining the existing predictive framework by optimizing data pre-processing and feature extraction to handle larger and more heterogeneous clinical datasets. The integration of medical ontologies such as ICD-11 or SNOMED CT can enhance the model's semantic understanding of clinical terms and improve interoperability with electronic health record (EHR) systems.

Additionally, incorporating transformer-based architectures such as BERT or Bio BERT could further enhance contextual representation of clinical narratives. Explainability mechanisms, including attention visualization, will be introduced to interpret model decisions and increase trust among healthcare professionals.

To improve system adaptability, the model can be extended to support multilingual datasets and real-time ICD code prediction within hospital information systems. Finally, the system's integration with clinical analytics modules can support early disease detection and personalized treatment recommendations, expanding its applicability beyond automated coding toward intelligent healthcare decision support.

REFERENCES

[1] Y. Abdullah R., A. S. Selvaraj, S. N., and H. R. T., "Chatbot for Healthcare System Using NLP and Python," Proc. 2022 IEEE Int. Conf. on Service Operations and Logistics, and Informatics (SOLI), pp. 1–6, 2022, doi: 10.1109/SOLI57430.2022.10294817.

[2] C.-C. Hsu, P.-C. Chang, and A. Chang, "Multi-Label Classification of ICD Coding Using Deep Learning," Proc. IEEE Conf. on Information Management and Technology, National Yunlin University of Science and Technology, Yunlin, Taiwan, 2021.

[3] Y. Chen and J. Ren, "Automatic ICD Code Assignment Utilizing Textual Descriptions and Hierarchical Structure of ICD Code," Proc. IEEE Int. Conf. on Data and Computer Science, Sun Yat-Sen University, Guangdong, China, 2021.

[4] P. Dohare, S. Johri, S. Priya, S. Singh, and S. Upadhyay, "Good Fellow: A Healthcare Chatbot System," Proc. IEEE Conf. on Electrical Engineering and Technology (DEI), Agra, India, 2022.

[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. Int. Conf. on Learning Representations (ICLR), 2013.

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] S. Reddy and J. Aggarwal, "A Survey on the Application of

Machine Learning Models in ICD Code Prediction," IEEE Access, vol. 10, pp. 12543–12556, 2022.

[8] R. Johnson and T. Zhang, "Supervised and Semi-Supervised Text Categorization Using LSTM for ICD Code Prediction,"

Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM), pp. 1012–1018, 2021.

[9] M. Singh, M. Mittal, P. Dewan, A. Kaur, G. Kaur, and A. Gupta, "Artificial Intelligence Chatbots in Healthcare," Proc. 2024 11th Int. Conf. on Computing for Sustainable Global Development (INDIACom), 2024.

[10] N. Goyal, R. Arora, and P. Kumar, "Comparative Analysis of Machine Learning Algorithms for Disease Prediction," IEEE Access, vol. 11, pp. 56345–56358, 2023.

[11] H. Luo, M. Ding, and Y. Wang, "Deep Neural Architectures for Clinical Text Understanding and Disease Coding," IEEE Trans. Neural Networks and Learning Systems, vol. 33, no. 7, pp. 3190–3204, 2022.

[12] K. Sharma and R. Bhattacharya, "A Hybrid NLP and Deep Learning Approach for Clinical Text Classification," IEEE J. of Biomedical Informatics, vol. 29, no. 3, pp. 112–118, 2023.

[13] D. Ramesh and C. Vardhan, "Disease Prediction Using Machine Learning Models in Healthcare," Proc. IEEE Int. Conf. on Computational Intelligence and Data Science (ICCIDS), pp. 1–6, 2021.

[14] P. Srivastava, A. Jain, and S. Kumar, "Medical Text Classification Using LSTM Networks and Word Embeddings," Proc. IEEE Region 10 Conf. (TENCON), pp. 1782–1787, 2022.

[15] L. Zhou, Y. Wang, and M. Zhang, "Explainable Deep Learning Models for Clinical Diagnosis Support," IEEE Access, vol. 12, pp. 27390–27402, 2024.