MEDICAL CODING

GUIDE:
MRS. LAKSHMI SURESH

ANNA PRIZE JOHNEY
AYSHA NAZRIN AFSAL
LEKSHMIPRIYA C M

CONTENT

01 INTRODUCTION

02 PROBLRM STATEMENT

03 OBJECTIVE

04 LITERATURE SURVEY

05 PROPOSED SYSTEM

06 CONCLUSION

07 REFERENCE

INTRODUCTION

- In the realm of modern healthcare, efficient and standardized medical coding is crucial for accurate communication, billing, and research.
- This project focuses on addressing challenges within the International Classification of Diseases (ICD-10) system, such as biased distributions and high variability.
- By exploring example-based methods and leveraging semantic features, we aim to improve the precision of medical coding.
- Purpose is strive to enhance the efficiency of healthcare systems and contribute to improved patient care outcomes.

PROBLEM STATEMENT

- Communication Medium is the predominant written texts such as clinical notes, electronic health records, medical reports.
- Challenge faced in Medical Coding are inefficiencies and inaccuracies due to the complexity of healthcare narratives and also risk of misinterpretations and manipulations for personal gain.
- These intricacies underscore the need for an advanced system to decipher and accurately translate medical text into standardized codes.
- Aiming to reduce errors, enhance accuracy, and ensure reliable healthcare documentation.

OBJECTIVE

- To formulate a user-friendly web application with a simple and efficient interface tailored to the needs of healthcare professionals.
- To develop a system proficient in identifying and assigning accurate medical codes, with a primary focus on emotions related to medical conditions, such as urgency, severity, and complexity.

LITERATURE SURVEY

Paper	Technology	Advantages	Disadvantages
Automatic medical code assignment via Deep learning approach for intelligent healthcare	Medic al topic mining, cross-textual attention mechanism, and auxiliary coding	• Achieves high accuracy • Significantly reduces coding time	• Data dependency • Black box nature
UMLS mapping and Word embeddings for ICD code assignment using the MIMIC-III intensive care database	Combination of Support Vector Machines (SVM) and FastText with Unified Medical Language System (UMLS) metathesaurus mappings	• Accurate • Efficient ICD coding, • Better diagnosis and treatment	• Developed and evaluated for the English language only.
Leveraging Semantics in WordNet to Facilitate the Computer-Assisted Coding of ICD-11	Semantic similarity-based approach to ICD-11 coding	• More explainable • More applicable to multiple languages	• Less accurate • Less efficient • More complex to implement
Extreme Multi-Label ICD Classification: Sensitivity to Hospital Service and Time	ElMo embeddings, attention mechanisms, data augmentation, ensemble methods	• Outperforms existing methods improve accuracy and efficiency of ICD coding	• Difficult to interpret predictions. • Only evaluated on a single dataset

PROPOSED SYSTEM

- The MEDCODE (Medical Coding Neural Network) is an innovative system that revolutionizes medical coding through advanced NLP techniques.
- Its dual sub-network architecture, combining BiLSTM and CNN components, enables comprehensive understanding of healthcare narratives and extraction of coding-specific features.
- The convolutional layers and pooling operations are tailored to capture both sentence and word embedding dimensions, enhancing feature extraction.
- The proposed VCPCNN architecture aims to overcome the limitations of traditional methods, leveraging deep learning advancements for comprehensive sentiment analysis in medical coding.
- The network's flexibility allows it to adapt to varying sentence structures and languages, making it a promising approach for nuanced sentiment classification in healthcare texts.
- The web application interface provides healthcare professionals with user-friendly access to input text and receive accurate code suggestions, streamlining coding processes and improving documentation accuracy.
- MEDCODE represents a transformative approach to medical coding, poised to enhance efficiency in healthcare documentation.

ARCHITECTURE

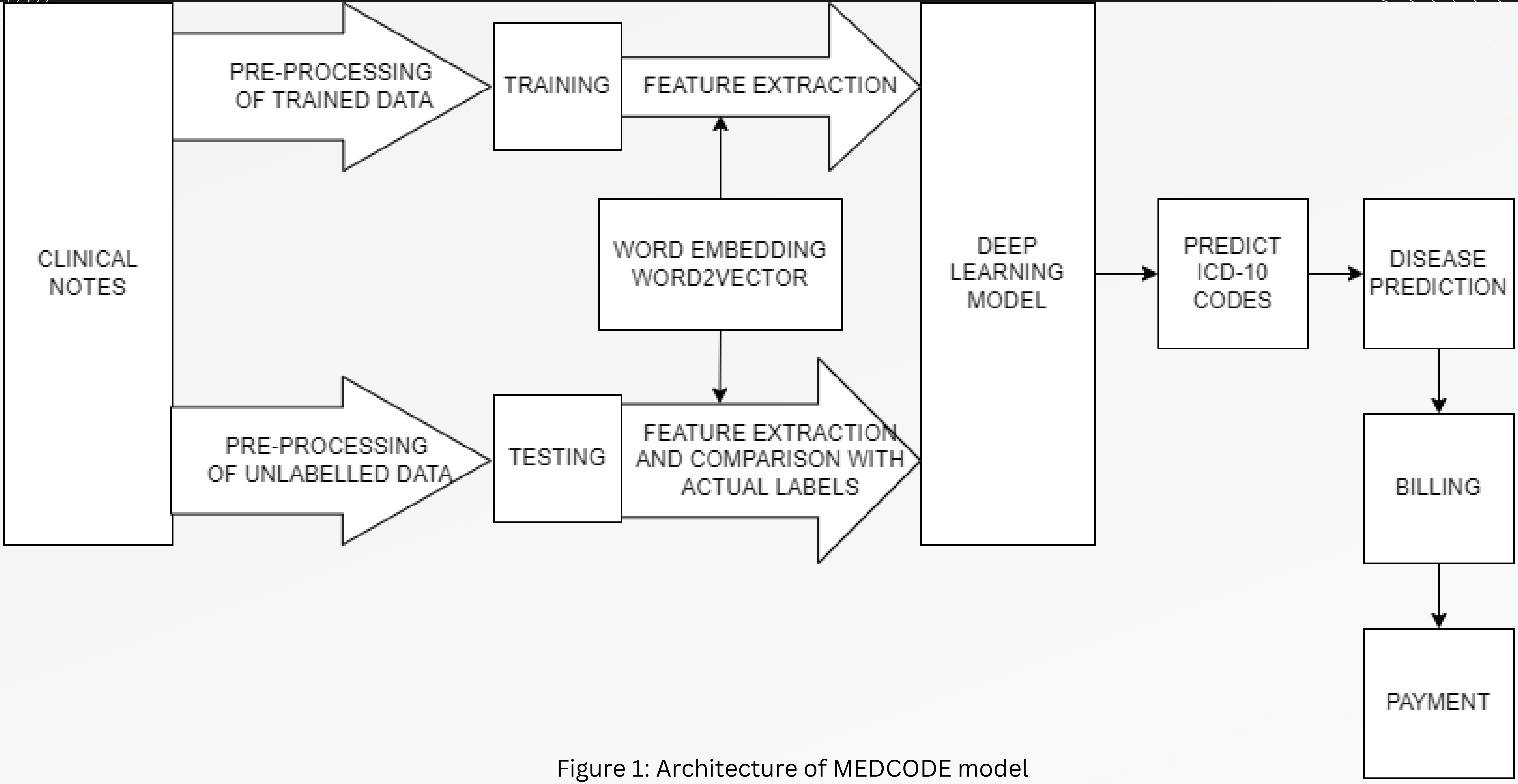

Figure 1: Architecture of MEDCODE model

MODULES

1. Data Preprocessing Module

- Data Preprocessing Module is crucial for ensuring that the input data is clean, formatted, and structured in a suitable manner for training and prediction tasks.

Text Cleaning and Formatting:

- One of the primary tasks within the Data Preprocessing Module is text cleaning and formatting.
- This step involves removing any irrelevant characters, punctuation, special symbols, and whitespace from the raw text data.
- By standardizing the text in this manner, the module aims to make it uniform and ready for further processing.
- Algorithms such as regular expressions (Regex) are commonly utilized for identifying and removing specific patterns or characters from the text.

Tokenization:

- Tokenization involves breaking down the text into smaller units, such as words or phrases.
- This process is essential for creating a structured representation of the text data that can be easily processed by the neural network.
- Various techniques, including whitespace-based tokenization, regular expression-based tokenization, or advanced tokenizers like those found in the NLTK (Natural Language Toolkit) library, are commonly employed for this purpose.

MODULES

Word Embedding:

- Once tokenized, the text data is transformed into word embeddings, which are dense vector representations of words in a continuous vector space.
- Word embedding captures the semantic relationships between words, enhancing the neural network's understanding of the text.
- Algorithms such as Word2Vec, GloVe (Global Vectors for Word Representation), and FastText are commonly used for generating word embeddings from large text corpora.

Data Augmentation (Optional):

- In some cases, data augmentation techniques may be applied to increase the diversity and size of the training data.
- Data augmentation techniques for text data may include adding synonyms or paraphrases, randomly inserting or deleting words, or applying transformations such as word shuffling or replacement.

Padding and Truncation:

- Finally, to ensure uniform input sizes across all data samples, padding or truncation may be applied to the text data.
- These operations adjust the length of the input sequences to meet the requirements of the neural network, which typically expects fixed-length input sequences.
- Padding involves adding zeros or a special token to the end of sequences, while truncation involves cutting off excess tokens.
- Libraries like TensorFlow or PyTorch provide convenient functions for handling variable-length sequences and performing padding and truncation operations.

MODULES

2. Bidirectional Long-Short Term Memory (BiLSTM) Sub-Network:

- The Bidirectional Long-Short Term Memory (BiLSTM) sub-network is a crucial component of the MEDCODE (Medical Coding Neural Network) system, designed to capture contextual and semantic nuances in healthcare narratives.
- It enhances the understanding of the sequential nature of medical text by processing data bidirectionally, ensuring that dependencies in both forward and backward directions are captured effectively.
- This sub-network is adept at learning long-term dependencies, making it well-suited for tasks involving sequential data such as text.

Forward and Backward Passes:

- During the forward pass, the input sequence is processed from the beginning to the end, with each LSTM unit updating its cell state and output based on the current input and the previous hidden state.
- During the backward pass, the input sequence is processed from the end to the beginning, allowing the network to capture dependencies in the reverse order.

$$p_t = p_t^f + p_t^b$$

p_t : Final probability vector of the network.

p_t^f :Probability vector from the forward LSTM network.

p_t^b : Probability vector from the backward LSTM network.

MODULES

3. Convolutional Neural Network (CNN) Sub-Network:

- Convolutional layers apply convolution operations to the input data using learnable filters (kernels). These filters slide over the input text data, extracting local patterns or features.
- Each filter captures different features or patterns, such as word combinations or phrases, through a process known as feature mapping.
- Convolutional layers are characterized by parameters such as filter size, stride, and padding, which determine the size and shape of the output feature maps.

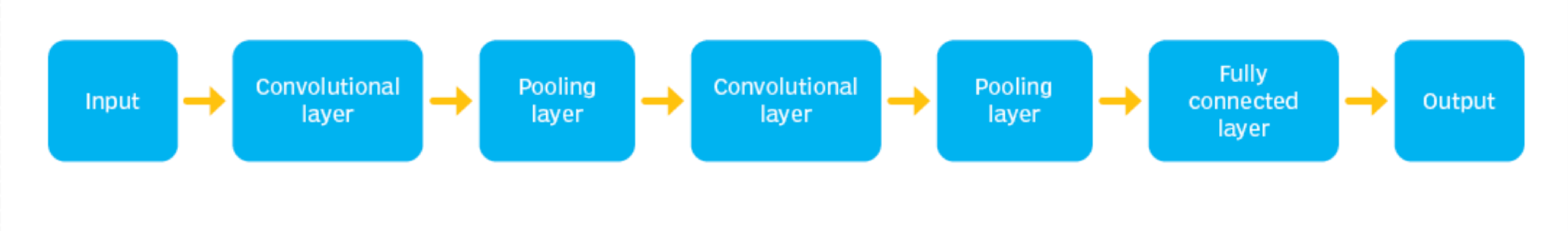

Figure 2: Structure of CNN

MODULES

4. Code Prediction Module:

- The Code Prediction Module in the MEDCODE (Medical Coding Neural Network) system is responsible for accurately predicting medical codes based on the features extracted from the BiLSTM and CNN sub-networks.
- This module integrates a Random Forest algorithm, which is a supervised learning algorithm capable of handling both classification and regression tasks.
- Random Forest is chosen for its robustness, ability to handle high-dimensional data, and capability to deal with non-linear relationships between features and target labels.

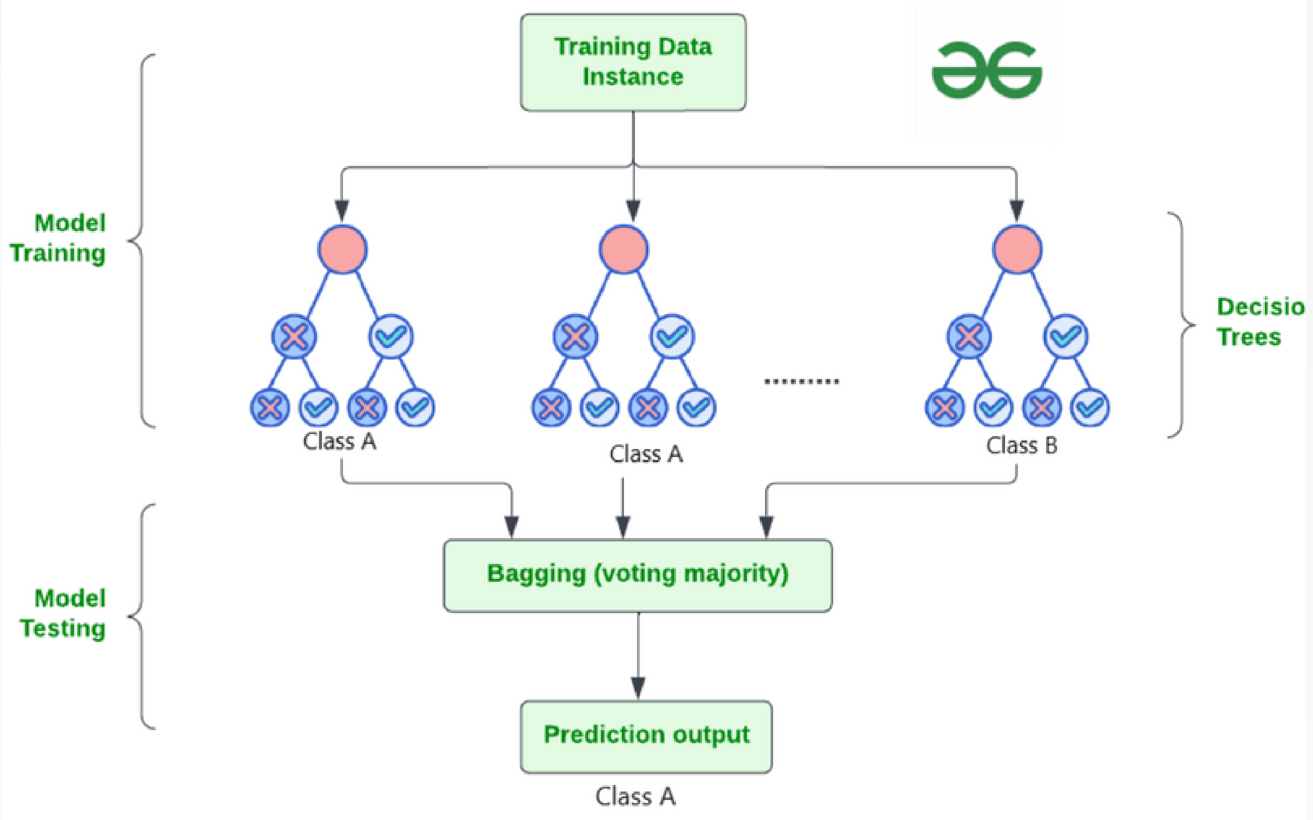

Figure 3: Working of Random Forest algorithm

Algorithm 1: The RF algorithm.

Training Phase :

Given :

- D : training set with n instances, p features, and target variable.
- K : number of classes in target variable.
- B : number of classifiers in RF.

Procedure :

For $b = 1$ to B

1. Generate bootstrapped sample D_b^* from the training set D.
2. Grow a tree using a random feature subset from bootstrapped sample D_b^*.

   For a given node t,

   (i) Randomly select $m \approx \sqrt{p}$ or $m \approx p/3$ feature.

   (ii) Find the best split features and cutpoints using the random feature subset.

   (iii) Send down the data using the best split features and cutpoints.

   Repeat (i)-(iii) until stopping rules are met.

3. Construct trained classifiers C_b.

Test Phase :

Aggregate the B trained classifiers using simple majority vote. For a test instance x, the predicted class label from classifiers C_B is given as :

$$C_B(x) = \text{argmax}_j \sum_{b=1}^{B} I(C_b(x) = j), \text{for } \ j = 1, \ldots, K$$

MODULES

5. Web Application Interface:

- The Web Application Interface of the MEDCODE system provides healthcare professionals with a user-friendly platform to input medical text data and receive automated suggestions for medical code assignment.
- It features an intuitive layout with an input text field for data entry and a section for displaying automated code suggestions.
- Users can review, select, and confirm the suggested codes, with options for feedback and error handling. The interface is designed for accessibility, compatibility across devices, and compliance with security and privacy regulations, ensuring a seamless and secure user experience.

6. Medicines and Diagnosis Prediction Module:

- The Medicines Prediction Module is a component within a healthcare system designed to predict and suggest appropriate medications for patients based on the result obtained after the prediction of the ICD-10 codes.
- In this module we use NLP algorithms like tokenization, stop words removal, lemmatization etc.

MODULES

7. Billing and Payment Module:

- Automating the billing and payment module using Python streamlines the process of generating bills, verifying insurance coverage, calculating charges, and processing payments.
- Python scripts integrate data from various sources, generate itemized bills, verify insurance details, and handle payment processing through APIs or payment gateways.
- Automating this module using Python can streamline administrative tasks, reduce errors, and improve efficiency.
- Different methods used are: Data Integration, Billing Generation, Insurance Verification, Calculation of Charges, Payment Processing etc.
- By automating the billing and payment module using Python, healthcare organizations can streamline administrative tasks, reduce manual errors, improve accuracy, and enhance overall efficiency in managing financial transactions related to medical services provided to patients.

CONCLUSION

- The "Medical Coding Analyzer," utilizing the Random Forest algorithm, proves invaluable in discerning and categorizing emotions in medical coding, tailored for the C-10 dataset.
- Its emotion detection feature enhances users' understanding of conveyed emotions, especially beneficial for professionals in gauging sentiment and optimizing communication.
- The application significantly contributes to fostering a positive online presence within the medical coding community.
- It stands as an essential resource offering valuable insights into emotional nuances, expected to remain a cornerstone for navigating emotional expression in medical coding discussions.

REFERENCES

[1] Almagro, M., Unanue, R. M., Fresno, V., & Montalvo, S. (2020). "ICD-10 Coding of Spanish Electronic Discharge Summaries: An Extreme Classification Problem." IEEE Access, 8, 100073–100083.

[2] Fei Teng, Zheng Ma, Jie Chen, Ming Xiao, and Lufei Huang, “Automatic medical code assignment via deep learning approach for intelligent healthcare”, Auckland University of Technology, May 26,2020 from IEEE Xplore.

[3] Liu, Z., Hu, Y., Wu, X., Mertes, G., Yang, Y., & Clifton, D. A. (2023). "Patient clustering for vital organ failure using ICD code with graph attention." IEEE Transactions on Biomedical Engineering, 70(8), 2329 2337.

[4] Henning Schafer , and Christoph M. Friedrich, “UMLS mapping and Word embeddings for ICD code assignment using the MIMIC-III intensive care database”, ©2019 IEEE.

[5] Donghua Chen, Runtong Zhang, “Leveraging Semantics in WordNet to Facilitate the Computer-Assisted Coding of ICD-11”, IEEE Journal of Biomedical and Health Informatics, 2019.

[6] ALBERTO BLANCO , ALICIA PÉREZ , AND ARANTZA CASILLAS, “Extreme Multi-Label ICD Classification: Sensitivity to Hospital Service and Time”, October 7, 2020.

[7] Siangchin, N., & Samanchuen, T. (2019). "Chatbot Implementation for ICD-10 Recommendation System." IEEE.

[8] Ahmed, A. S., & Saifuddin, K. M. (2017). "A Simplistic, Effective, and Adaptive Approach towards Classifying Medical Records according to ICD-10 using Machine Learning for Efficient Statistics." IEEE.

THANK YOU